

# Analyse graphischer Integrationsmethoden

## INAUGURALDISSERTATION

zur Erlangung der philosophischen Doktorwürde, vorgelegt der  
Philosophisch-naturwissenschaftlichen Fakultät  
der Universität Basel

von

**WALTER GAUTSCHI**

aus Reinach (Kt. Aargau) und Basel

Basel, 1953/54

E. Birkhäuser & Cie. AG.



# Analyse graphischer Integrationsmethoden

## INAUGURALDISSERTATION

zur Erlangung der philosophischen Doktorwürde, vorgelegt der  
Philosophisch-naturwissenschaftlichen Fakultät  
der Universität Basel

von

**WALTER GAUTSCHI**

aus Reinach (Kt. Aargau) und Basel

Basel, 1953/54

E. Birkhäuser & Cie. AG.

Genehmigt von der Philosophisch-naturwissenschaftlichen  
Fakultät auf Antrag der

Herren Professoren Dr. A. Ostrowski und Dr. A. Speiser.

Basel, den 2. Juli 1953

Prof. Dr. P. Huber (Dekan)



Separatabdruck aus den *Verhandlungen der Naturforschenden Gesellschaft in Basel*

Band 64 (1953), S. 401–435

Band 65 (1954), S. 49– 66

# Inhalt

---

## *Fehlerabschätzungen für die graphischen Integrationsverfahren von Grammel und Meissner-Ludwig*

|                      |     |
|----------------------|-----|
| Einleitung . . . . . | 401 |
|----------------------|-----|

### Kapitel I

#### Die Integrationsverfahren von Grammel und Meissner-Ludwig im Falle der Quadratur

|                                                      |     |
|------------------------------------------------------|-----|
| § 1. Das Grammelsche Orthopolarenverfahren . . . . . | 403 |
| § 2. Das Meissnersche Linienbildverfahren . . . . .  | 407 |
| § 3. Die Quadraturfehler . . . . .                   | 410 |

### Kapitel II

#### Die Integrationsverfahren im Falle von Differentialgleichungen $n$ -ter Ordnung und rekursive Abschätzung des Fehlers

|                                                          |     |
|----------------------------------------------------------|-----|
| § 4. Bezeichnungen und Voraussetzungen . . . . .         | 411 |
| § 5. Roh- und Sehnenverfahren . . . . .                  | 414 |
| § 6. Tangenten- und Sehnen-Tangenten-Verfahren . . . . . | 417 |

### Kapitel III

#### Independente Fehlerabschätzung im Falle von Differentialgleichungen $n$ -ter Ordnung

|                                                                        |     |
|------------------------------------------------------------------------|-----|
| § 7. Ein System von linearen Differenzengleichungen . . . . .          | 422 |
| § 8. Das Prinzip der Fehlerabschätzung für allgemeines $n$ . . . . .   | 426 |
| § 9. Explizite Fehlerschranken für $n = 1$ und $n = 2$ . . . . .       | 430 |
| § 10. Asymptotischer Vergleich der drei Halbschrittverfahren . . . . . | 433 |

## *Über die zeichnerischen Ungenauigkeiten und die zweckmässige Bemessung der Schritt-länge beim graphischen Integrationsverfahren von Meissner-Ludwig*

|                                                  |    |
|--------------------------------------------------|----|
| § 1. Einleitung . . . . .                        | 49 |
| § 2. Bezeichnungen und Voraussetzungen . . . . . | 54 |
| § 3. Diskussion der technischen Fehler . . . . . | 56 |
| § 4. Beispiel . . . . .                          | 62 |

---



Digitized by the Internet Archive  
in 2026

# Fehlerabschätzungen für die graphischen Integrationsverfahren von Grammel und Meissner-Ludwig

Mit 10 Abbildungen

Von Walter Gautschi, Basel

## Einleitung

In mehreren Arbeiten hat R. LUDWIG (1951a, b, c)<sup>1)</sup> die bekannten graphischen Integrationsverfahren von MEISSNER (1931/32)<sup>2)</sup> und GRAMMEL (1939)<sup>2)</sup> zum erstenmal analytisch untersucht und ausserdem eine sehr bemerkenswerte Neufassung des Meissnerschen Linienbildverfahrens gegeben. Diese Neufassung ist durch eine weitgehende Analogie mit dem Grammelschen Verfahren gekennzeichnet. Insbesondere ergeben sich einheitliche Konstruktionen für Differentialgleichungen von beliebiger Ordnung  $n$ , und der Fall  $n = 1$  braucht nicht mehr, wie bisher<sup>3)</sup>, gesondert behandelt zu werden. Die Genauigkeit dieses Meissner-Ludwigschen Verfahrens ist dieselbe wie diejenige des Grammelschen Verfahrens. Gegenüber dem letzteren zeichnet es sich aber dadurch aus, dass es frei von „Singularitäten“ ist.

Für beide Verfahren macht LUDWIG (1951 b, c) auch Angaben über Fehler und Fehlerfortpflanzung, indem er einen klassischen Abschätzungssatz aus der Theorie der Differentialgleichungen verwendet. Dieser Satz und ähnliche Sätze sind zu diesem Zweck schon wiederholt [TOLLMIEHN (1938), FRICKE (1949)], wenn auch nicht ganz einwandfrei, bei numerischen Methoden angewandt worden<sup>4)</sup>.

Auch die Abschätzung, die LUDWIG gibt, ist, vor allem aus zwei Gründen, ungenau. Einmal wird das Maximum  $\theta$  des Inte-

<sup>1)</sup> Die Jahreszahlen in Klammern verweisen auf das Literaturverzeichnis am Schluss der Arbeit (S. 434).

<sup>2)</sup> Vgl. auch KAMKE (1944), S. 171 ff.

<sup>3)</sup> Vgl. DOETSCH (1921), SCHULZ (1941).

<sup>4)</sup> Zur Kritik der Abschätzungen in TOLLMIEHN (1938) und FRICKE (1949) vgl. WEISSINGER (1950).

gralmittelwertes (11) in LUDWIG (1951 b, S. 170) ungenügenderweise nur für diskrete Werte von  $x$  abgeschätzt (S. 171). Andererseits ist die Funktion  $\bar{y}(x)$ , ausser an den Stellen  $x_n$ , nicht nur „nicht bekannt“, sondern wird überhaupt nicht definiert. Da die Quadraturfehler  $R_n$ , die in den Fehlerschranken auftreten, Ableitungen von  $\bar{y}$  enthalten, haben die Grössen  $R_n$  und damit auch die Schranken strenggenommen keinen Sinn.

Gegenstand dieser Arbeit wird sein, exakte Fehlerschranken für die Grammelschen und die Meissner-Ludwigschen Verfahren herzuleiten. Insbesondere soll auch versucht werden, die Schranken den jeweils vorliegenden Differentialgleichungen anzupassen.

Wir verwenden dabei eine Abschätzungstechnik, die auf von MISES (1930) zurückgeht, der zum erstenmal im Falle des Adamschen Verfahrens für Differentialgleichungen erster Ordnung die Fortpflanzung des Fehlers durch eine Differenzenungleichung beschrieben hat. Diese Methode wird sinngemäss auf Differentialgleichungen  $n$ -ter Ordnung ausgedehnt werden und führt auf ein *System* von  $n$  Differenzenungleichungen. Die beiden ersten Kapitel werden der Herleitung dieser Systeme gewidmet sein, die dann im dritten Kapitel zu praktisch brauchbaren Fehlerabschätzungen verarbeitet werden.

In den §§ 1 und 2 des ersten Kapitels werden die verschiedenen Varianten des Grammelschen und des Meissner-Ludwigschen Verfahrens im Falle von Quadraturen systematisch dargestellt und auf die enge Verwandtschaft der beiden Verfahrensgruppen hingewiesen (Nr. 7). In § 3 werden die entsprechenden Quadraturfehler angegeben, die im wesentlichen auf Restglieder bekannter Quadraturformeln hinauslaufen. Die Ausdrücke für die Quadraturfehler werden im zweiten Kapitel bei der Fehlerabschätzung für Differentialgleichungen Verwendung finden.

Die Voraussetzungen, unter denen die Abschätzung durchgeführt wird, sind in § 4 zusammengestellt. Es wird die nicht-negative Matrix  $K$  eingeführt, die charakteristische Daten der vorgegebenen Differentialgleichung enthält. Die Eigenwerte dieser Matrix, insbesondere die Maximalwurzel, werden nicht nur für die Fortpflanzung des Fehlers, sondern auch für die Konvergenz gewisser Iterationen eine massgebende Rolle spielen. Im Anschluss an die Beschreibung der Integrationsverfahren werden in den §§ 5 und 6 für die Fehler die entsprechenden Differenzenungleichungssysteme hergeleitet, die sich in Matrizenform mit Hilfe der Matrix  $K$  übersichtlich schreiben lassen.

Zur weiteren Diskussion dieser Systeme in Kapitel III wird ein entsprechendes Differenzenungleichungssystem betrachtet, deren

allgemeine Lösung in § 7 angegeben wird. Mit Hilfe dieser Lösung werden in § 8 im Falle von Differentialgleichungen beliebiger Ordnung  $n$  zwei Abschätzungssätze angegeben, die zu Fehlerschranken verschiedener Güte führen (Satz I und II). Für die praktisch wichtigen Fälle  $n = 1$ ,  $n = 2$  werden in § 9 die Schranken soweit explizit hingeschrieben, dass sie unmittelbar auf sämtliche Integrationsvarianten angewendet werden können. Schliesslich wird in § 10 das asymptotische Verhalten der Fehlerschranken für die verschiedenen Varianten miteinander verglichen.

Die bei graphischen Methoden besonders schwerwiegenden *zeichnerischen Fehler* werden an anderer Stelle untersucht werden, wo dann auch ein Beispiel einer vollständigen Fehlerabschätzung durchgeführt wird.

Meinem hochverehrten Lehrer, Herrn Prof. Dr. A. OSTROWSKI, bin ich für sein Interesse, das er dem Entstehen dieser Arbeit entgegengebracht hat, und für manche wertvolle Anregungen zu tiefstem Dank verpflichtet.

## Kapitel I

### Die Integrationsverfahren von Grammel und Meissner-Ludwig im Falle der Quadratur

#### § 1. Das Grammelsche Orthopolarenverfahren

1. Die graphische Integration nach GRAMMEL (1939) beruht auf einer eigentümlichen Art, reelle Funktionen  $y(x)$  eines reellen Argumentes  $x$  darzustellen. Als Bild einer solchen Funktion verwendet man die Kurve<sup>1)</sup>

$$G(x) = \frac{1}{y(x)} e^{ix} \quad [y(x) \neq 0]. \quad (1)$$

Wir nennen  $G(x)$  die *Grammelsche Polare*<sup>2)</sup> der Funktion  $y(x)$ . Ähnlich definiert man die Bilder  $G_v(x)$  der Ableitungen  $y^{(v)}(x)$ , die sogenannten  *$v$ -ten Orthopolaren* oder auch die  *$v$ -ten Grammelschen Polaren* der Funktion  $y(x)$ , durch

$$G_v(x) = \frac{i^v}{y^{(v)}(x)} e^{ix} \quad [y^{(v)}(x) \neq 0; v = 0, 1, 2, \dots], \quad G_0(x) = G(x). \quad (2)$$

<sup>1)</sup> Wenn hier und im folgenden kurz von der „Kurve  $G(x)$ “ gesprochen wird, so ist damit die Gesamtheit der Punkte verstanden, die den komplexen Zahlen  $G(x)$  in (1) zugeordnet sind.

<sup>2)</sup> Die Grammelsche Polare wird in der ganzen Arbeit durchweg mit einer (noch zu definierenden) Funktion  $y$  gebildet, so dass wir in der Bezeichnung (1) dieser Polaren die Funktion  $y$  nicht ausdrücklich vermerken wollen.

Wegen

$$\frac{d}{dx} G(x) = \frac{y'}{y} [G_1(x) - G(x)] \quad (3)$$

hat der Tangentenvektor der Polaren  $G(x)$  in einem Punkt  $P$  stets die Richtung  $PP'$  zum entsprechenden Punkt  $P'$  der ersten Orthopolaren oder die Richtung  $P'P$ , je nachdem, ob  $y y' > 0$  bzw.  $y y' < 0$  ist (vgl. Abb. 1).

Ein entsprechender Sachverhalt besteht natürlich auch zwischen den höheren Orthopolaren, worauf geradezu die Möglichkeit beruhen wird, die Integration von Differentialgleichungen beliebig hoher Ordnung in sehr natürlicher Weise durchzuführen.

2. Die Quadraturaufgabe sei in diesem Zusammenhang in Form einer Differentialgleichung

$$y'(x) = f(x) \quad (4)$$

mit vorgegebener Funktion  $f(x)$  und einer Anfangsbedingung

$$y(x_0) = y_0 \quad (4a)$$

gegeben. Es besteht die Aufgabe, zur bekannten ersten Orthopolaren

$$G_1(x) = \frac{i}{f(x)} e^{ix}$$

der gesuchten Funktion  $y(x)$  die Polare  $G(x)$ , deren Anfangspunkt  $G_0 = (1/y_0) e^{ix_0}$  man kennt, annähernd zu bestimmen.

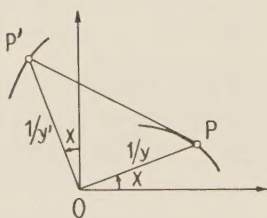


Abb. 1. Grammelsche Polaren

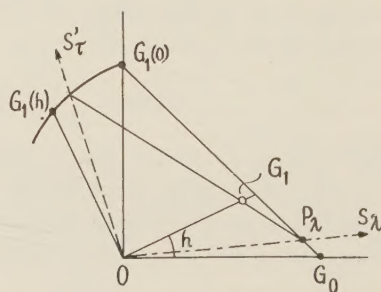


Abb. 2. Teilschrittverfahren

Wir wählen eine geeignete Schrittlänge  $h > 0$  und definieren eine Folge von äquidistanten Argumentsstellen

$$x_\varrho = x_0 + \varrho h \quad (\varrho = 0, 1, 2, \dots, N), \quad h > 0, \quad (5)$$

in denen Näherungen  $y_\varrho$  der gesuchten Funktionswerte  $y(x_\varrho)$  bestimmt werden sollen. Generell bezeichnen wir Näherungspunkte für  $G_\nu(x_\varrho)$  mit  $G_\varrho^{(\nu)}$  und setzen insbesondere

$$G_\varrho = \frac{1}{y_\varrho} e^{ix_\varrho}.$$

Ferner bezeichnen wir die radialen Strahlen, die zu den Winkeln  $x_\varrho$  gehören, mit  $s_\varrho$ , und allgemein werden wir unter  $s_\varrho^{(\nu)}$  die Strahlen verstehen, die aus  $s_\varrho$  durch Drehung im positiven Sinn um den Winkel  $\nu\pi/2$  hervorgehen.

Am einfachsten erhält man Näherungspunkte  $G_\varrho$ , auf Grund der Eigenschaft (3), nach folgender Vorschrift. Man konstruiere, beginnend mit  $\varrho = 0$ , den Punkt  $G_{\varrho+1}$  aus dem bereits bekannten Punkt  $G_\varrho$  als den Schnittpunkt der Verbindungsgeraden  $G_1(x_\varrho)G_\varrho$  mit dem Strahl  $s_{\varrho+1}^{(1)}$ . Der analytische Ausdruck dieser Vorschrift ist durch die Relationen

$$y_{\varrho+1} = y_\varrho \cos h + f_\varrho \sin h, \quad f_\varrho = f(x_\varrho) \quad (\varrho = 0, 1, 2, \dots, N-1) \quad (\text{RV}) \quad (6)$$

gekennzeichnet. Es gilt dann näherungsweise

$$y_N - y_0 \approx \int_{x_0}^{x_N} f(x) dx.$$

Wir nennen (6) das *Rohverfahren* (RV) im Falle der Quadratur, weil in (6) im wesentlichen (das heisst bis auf Glieder zweiter Ordnung in  $h$ ) die noch sehr rohe Quadraturformel  $y_{\varrho+1} - y_\varrho + hf_\varrho$  steckt.

3. Entsprechend dem Umstand, dass durch Verwendung von Zwischenwerten  $f_{\varrho+\lambda} = f(x_\varrho + \lambda h)$ ,  $0 < \lambda < 1$ , der Funktion  $f(x)$  die Genauigkeit der Quadraturformeln wesentlich erhöht werden kann, hat man auch hier nach Konstruktionen gesucht, die geeignete Zwischenwerte  $x_{\varrho+\lambda} = x_\varrho + \lambda h$  des Arguments heranziehen.

Bezeichnet  $P_{\varrho+\lambda}$  den Schnittpunkt des Strahles  $s_{\varrho+\lambda}$  mit der Verbindungsgeraden  $G_1(x_\varrho)G_\varrho$ , so liegt es nahe, den Näherungspunkt  $G_{\varrho+1}$  als Schnittpunkt der Verbindungsgeraden  $G_1(x_{\varrho+\tau})P_{\varrho+\lambda}$  mit  $s_{\varrho+1}$  anzusetzen.  $\lambda, \tau$  sind dabei gewisse Zahlen zwischen 0 und 1. (In Abb. 2 ist der Fall  $\varrho = x_0 = 0$  gezeichnet.) Wir nennen eine solche Konstruktion allgemein ein *Teilschrittverfahren* und sprechen insbesondere von einem *Halbschrittverfahren*, wenn eine

<sup>1)</sup> Im Falle  $f(x_0) = 0$  ergibt sich sinngemäss  $G_{\varrho+1}$  als Schnittpunkt der Parallelen zu  $s_\varrho'$  durch  $G_\varrho$  mit  $s_{\varrho+1}$ .

der Zahlen  $\lambda$  oder  $\tau$  gleich  $1/2$  und  $\tau \neq 0$  ist. Der Grenzfall  $\tau = 0$  gibt, unabhängig vom Wert von  $\lambda$ , das Rohverfahren.

Zur Festlegung der Konstanten  $\lambda, \tau$  wird man verlangen, dass die Entwicklung des ersten Näherungswertes  $y_1$  nach Potenzen von  $h$  mit der Entwicklung des exakten Wertes

$$y(x_1) = y_0 + \int_{x_0}^{x_0+h} f(x) dx$$

bis zu Gliedern einer gewissen Ordnung übereinstimmt. Man nennt den maximalen Exponenten von  $h$ , bei dem noch Übereinstimmung gilt, die *Ordnung* des Verfahrens.

Nun berechnet man  $y_1$  in Abhängigkeit von  $y_0, f_0, f_\tau$  und den Konstanten  $\lambda, \tau$  leicht zu

$$y_1 = \frac{1}{\cos(\lambda - \tau)h} [y_0 \cos \lambda h \cos(1 - \tau)h + f_0 \cos(1 - \tau)h \sin \lambda h + f_\tau \sin(1 - \lambda)h]. \quad (7)$$

Entwickelt man den Ausdruck in der eckigen Klammer nach Potenzen von  $h$ , so erhält man

$$\begin{aligned} y_1 \cos(\lambda - \tau)h &= y_0 + f_0 h + \frac{1}{2!} \{2\tau(1 - \lambda)f'_0 - [\lambda^2 + (1 - \tau)^2]y_0\}h^2 + \\ &+ \frac{1}{3!} \{3\tau^2(1 - \lambda)f''_0 - [\lambda^3 + 3\lambda(1 - \tau)^2 + (1 - \lambda)^3]f_0\}h^3 + \\ &+ \frac{1}{4!} \{4\tau^3(1 - \lambda)f'''_0 - 4\tau(1 - \lambda)^3f'_0 + \\ &- [\lambda^4 + (1 - \tau)^4 + 6\lambda^2(1 - \tau)^2]y_0\}h^4 + \dots \end{aligned}$$

Andererseits gilt

$$\begin{aligned} y(x_1) \cos(\lambda - \tau)h &= \left(y_0 + f_0 h + \frac{f'_0}{2!}h^2 + \frac{f''_0}{3!}h^3 + \frac{f'''_0}{4!}h^4 + \dots\right) \left(1 - \frac{(\lambda - \tau)^2}{2!}h^2 + \frac{(\lambda - \tau)^4}{4!}h^4 - \dots\right) = \\ &= y_0 + f_0 h + \frac{1}{2!} [f'_0 - (\lambda - \tau)^2 y_0] h^2 + \frac{1}{3!} [f''_0 - 3(\lambda - \tau)^2 f_0] h^3 + \\ &+ \frac{1}{4!} [f'''_0 - 6(\lambda - \tau)^2 f'_0 + (\lambda - \tau)^4 y_0] h^4 + \dots, \end{aligned}$$

und man sieht, dass das Teilschrittverfahren genau dann von zweiter Ordnung ist, wenn

$$\tau(1 - \lambda) = \frac{1}{2} \quad (8)$$

ist.

Ein Verfahren dritter Ordnung lässt sich noch erreichen, wenn man  $\lambda = 1/4$ ,  $\tau = 2/3$  setzt. Dies ist insofern bemerkenswert, als die Ordnung über 2 hinaus *direkt* erhöht werden kann, das heisst, ohne zwei verschiedene Verfahren miteinander zu kombinieren. Das Verfahren dritter Ordnung ist im Falle der Quadratur praktisch gut anwendbar. Für Differentialgleichungen eignet es sich weniger gut.

4. Seit GRAMMEL (1939) haben sich insbesondere zwei Halbschrittverfahren eingebürgert, das sogenannte *Sehnenverfahren* (SV) mit  $\lambda = 0$ ,  $\tau = 1/2$  und das sogenannte *Tangentenverfahren* (TV) mit  $\lambda = 1/2$ ,  $\tau = 1$ . Die analytischen Ausdrücke ergeben sich aus (7) im Falle des Sehnenverfahrens zu

$$y_{\varrho+1} = y_{\varrho} + 2f_{\varrho+\frac{1}{2}} \sin \frac{h}{2} \quad (\varrho = 0, 1, 2, \dots, N-1) \quad (\text{SV}) \quad (9)$$

und im Falle des Tangentenverfahrens zu

$$y_{\varrho+1} = y_{\varrho} + (f_{\varrho} + f_{\varrho+1}) \operatorname{tg} \frac{h}{2} \quad (\varrho = 0, 1, 2, \dots, N-1). \quad (\text{TV}) \quad (10)$$

In (9) und (10) stecken bis auf Glieder zweiter Ordnung in  $h$  die MacLaurinsche Quadraturformel bzw. die Trapezregel [LUDWIG (1951a)].

LUDWIG<sup>1)</sup> gibt ein unter Umständen gut brauchbares Verfahren vierter Ordnung an, indem er (9) und (10) zu

$$y_{\varrho+1} = \frac{2}{3} y_{\varrho+1}^{(\text{SV})} + \frac{1}{3} y_{\varrho+1}^{(\text{TV})}$$

kombiniert:

$$y_{\varrho+1} = y_{\varrho} + \frac{4}{3} f_{\varrho+\frac{1}{2}} \sin \frac{h}{2} + \frac{1}{3} (f_{\varrho} + f_{\varrho+1}) \operatorname{tg} \frac{h}{2} \quad (\varrho = 0, 1, 2, \dots, N-1). \quad (\text{STV}) \quad (11)$$

Man nennt (11) nach LUDWIG das *Sehnen-Tangenten-Verfahren*, das bis auf Glieder zweiter Ordnung in  $h$  mit der Simpsonschen Regel identisch ist.

## § 2. Das Meissnersche Linienbildverfahren

5. Im Unterschied zu GRAMMEL benutzt MEISSNER (1931/32) als Bild einer Funktion  $y(x)$  die Kurve<sup>2)</sup>

$$M(x) = y(x) e^{ix},$$

die wir als *Meissnersche Polare* der Funktion  $y(x)$  bezeichnen

<sup>1)</sup> LUDWIG (1951a), S. 126.

<sup>2)</sup> Das in den Fussnoten <sup>1)</sup> und <sup>2)</sup>, S. 403, über  $G(x)$  Gesagte gilt auch für  $M(x)$ .

werden. Ausserdem definiert MEISSNER das sogenannte *Linienbild*  $\mathfrak{M}(x)$  der Funktion  $y(x)$  als die Kurve

$$\mathfrak{M}(x) = [y(x) + i y'(x)] e^{ix}.$$

Analog zu (2) werden die  $\nu$ -ten *Polaren* und  $\nu$ -ten *Linienbilder* durch

$$M_\nu(x) = i^\nu y^{(\nu)}(x) e^{ix}$$

$$\text{bzw. } \mathfrak{M}_\nu(x) = i^\nu [y^{(\nu)}(x) + i y^{(\nu+1)}(x)] e^{ix} = M_\nu(x) + M_{\nu+1}(x)$$

eingeführt. Hier gilt nun

$$\frac{d}{dx} \mathfrak{M}(x) = i [y(x) + y''(x)] e^{ix}. \quad (12)$$

Sind also  $P$  und  $\mathfrak{P}$  Punkte der Polaren  $M(x)$  bzw. des Linienbildes  $\mathfrak{M}(x)$ , die zum gleichen Argument  $x$  gehören, so hat der *Tangentenvektor des Linienbildes  $\mathfrak{M}(x)$  im Punkte  $\mathfrak{P}$  die Richtung  $P\mathfrak{P}$  oder  $\mathfrak{P}P$* , je nachdem, ob an der Stelle  $x$  die Funktion  $y + y'' > 0$  bzw.  $< 0$  ist (vgl. Abb. 3). Entsprechendes gilt natürlich auch für die höheren Polaren und Linienbilder<sup>1)</sup>.

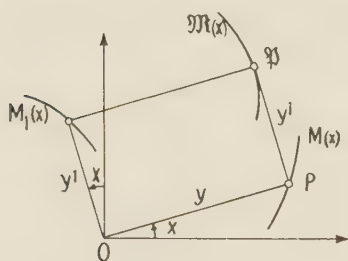


Abb. 3. Meissnersche Polaren und Linienbild

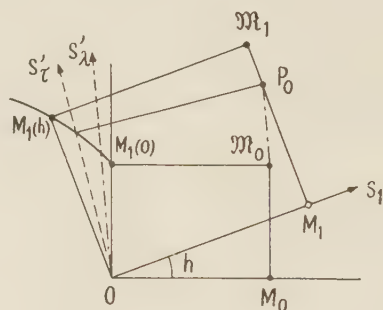


Abb. 4. Teilschrittverfahren für Linienbilder

6. Die Quadraturaufgabe sei wiederum durch (4) und (4a) formuliert. Dann kann offenbar die erste Meissnersche Polare  $M_1(x) = i f(x) e^{ix}$  der gesuchten Funktion  $y(x)$  gezeichnet werden, und ebenso der Anfangspunkt  $M_0 = y_0 e^{ix_0}$  der gesuchten Polaren  $M(x)$ . Um weitere Näherungspunkte  $M_2$  von  $M(x_0)$  für die Werte (5) des Argumentes  $x$  zu finden, wird das Linienbild, auf Grund der Eigenschaft (12), gewissermassen als Vermittler

<sup>1)</sup> Über weitere Eigenschaften des Linienbildes sei auf MEISSNER (1931–32) und LUDWIG (1951a) verwiesen.

herangezogen. Und zwar definieren wir gleich das allgemeine *Teilschrittverfahren für Linienbilder*:

Wir bezeichnen mit  $\mathfrak{M}_\varrho$  den Punkt  $(y_\varrho + i f_\varrho) e^{ix_\varrho}$ , der offenbar als Näherungspunkt des Linienbildes  $\mathfrak{M}(x)$  an der Stelle  $x_\varrho$  aufgefasst werden kann, und verstehen unter  $s_\varrho^{(\nu)}$  die in Nr. 2 eingeführten Strahlen. Sind  $\lambda, \tau$  zwei Zahlen zwischen 0 und 1, so ziehen wir, um  $M_{\varrho+1}$  aus  $M_\varrho$ ,  $\mathfrak{M}_\varrho$  und der Polaren  $M_1(x)$  zu konstruieren, vom Punkt  $M_1(x_{\varrho+\tau})$  aus eine Gerade senkrecht zu  $s'_{\varrho+\tau}$  und bringen sie zum Schnitt mit dem Strahl, der in  $\mathfrak{M}_\varrho$  in Richtung von  $s'_{\varrho+\lambda}$  angeheftet ist. Dieser Schnittpunkt sei  $P_\varrho$ , der noch von den Konstanten  $\lambda, \tau$  abhängt.  $M_{\varrho+1}$  ist dann der Fusspunkt des Lotes von  $P_\varrho$  auf  $s_{\varrho+1}$ . Mit  $M_{\varrho+1}$  ist natürlich auch die neue Näherung  $\mathfrak{M}_{\varrho+1}$  des Linienbildes durch

$$\mathfrak{M}_{\varrho+1} = (y_{\varrho+1} + i f_{\varrho+1}) e^{ix_{\varrho+1}}$$

festgelegt. (Vgl. Abb. 4, die für den Fall  $\varrho = x_0 = 0$  gezeichnet ist.)

7. Man verifiziert leicht die merkwürdige Tatsache, dass *der analytische Ausdruck für den ersten Näherungswert  $y_1$  des Meissnerschen Teilschrittverfahrens genau mit dem Ausdruck (7) für den entsprechenden Näherungswert des Grammelschen Teilschrittverfahrens übereinstimmt*. Daraus folgt unmittelbar, dass auch für das Meissnersche Teilschrittverfahren die Bedingung (8) notwendig und hinreichend ist, damit dieses Verfahren von zweiter Ordnung ist. Insbesondere spricht man nach LUDWIG (1951a) im Falle  $\lambda = 0$ ,  $\tau = 1/2$  vom *Tangentenverfahren* und im Falle  $\lambda = 1/2$ ,  $\tau = 1$  vom *Sehnenverfahren* für Linienbilder. Die analytische Formulierung des Tangentenverfahrens ist daher (9), diejenige des Sehnenverfahrens (10), also gerade umgekehrt wie im Falle der entsprechenden Grammelschen Verfahren. Die lineare Kombination

$$y_{\varrho+1} = \frac{2}{3} y_{\varrho+1}^{(\text{TV})} + \frac{1}{3} y_{\varrho+1}^{(\text{SV})}$$

liefert das *Sehnen-Tangenten-Verfahren* (11), das von vierter Ordnung ist. Schliesslich bezeichnen wir als *Rohverfahren* das Teilschrittverfahren mit  $\tau = 0$ .

Die gemeinsame analytische Fassung der Grammelschen und Meissnerschen Verfahren erlaubt, Fehlerabschätzungen zu geben, die in beiden Fällen gleichzeitig gültig sind. Ausserdem ist die Analogie zwischen beiden Verfahrensgruppen derart eng, dass wir die Übertragung der verschiedenen Halbschrittverfahren auf Differentialgleichungen explizit nur für den Fall der Grammel-

schen Verfahren ausführen werden. Die entsprechenden Konstruktionen für Linienbilder gehen daraus unmittelbar hervor. Auch die Bezeichnungen Sehnens- und Tangentenverfahren sind, falls nichts anderes gesagt wird, fortan stets im Grammelschen Sinne zu verstehen, das heisst gemäss (9) bzw. (10).

### § 3. Die Quadraturfehler

8. Zur Bestimmung der Quadraturfehler der in den Nrn. 2 und 4 beschriebenen Integrationsverfahren genügt es, den Fall  $N = 1$  zu betrachten. Zugleich nehmen wir, der einfachen Schreibweise halber,  $x_0 = 0$  an. Im Falle des Rohverfahrens gilt dann nach (6) für den Fehler  $R$

$$R = y_1 - y(x_1) = y_0 \cos h + f_0 \sin h - y_0 - \int_0^h f(x) dx = \\ = h f_0 - \int_0^h f(x) dx + y_0 (\cos h - 1) + f_0 (\sin h - h).$$

Durch Taylor-Entwicklung der drei Summanden rechter Hand erhält man daraus

$$R = -\frac{h^2}{2} \left\{ f'(\theta h) + y_0 \cos \theta_1 h + \frac{h}{3} f_0 \cos \theta_2 h \right\}. \quad (\text{RV}) \quad (13)$$

$\theta, \theta_1, \theta_2, \dots$  sind hier und im folgenden gewisse Zahlen zwischen 0 und 1.

In den übrigen Fällen lassen sich die Quadraturfehler ohne weiteres auf die geläufigen Restausdrücke bekannter numerischer Quadraturformeln zurückführen. So gilt nach (9) für das Sehnensverfahren

$$R = 2 f_{\frac{1}{2}} \sin \frac{h}{2} - \int_0^h f dx = R_M + f_{\frac{1}{2}} \left( 2 \sin \frac{h}{2} - h \right),$$

wo  $R_M$  das Restglied

$$R_M = h f_{\frac{1}{2}} - \int_0^h f dx = -\frac{h^3}{24} f''(\theta h)$$

der MacLaurinschen Regel ist. Wegen

$$\vartheta(h) \equiv 2 \sin \frac{h}{2} - h = -\frac{h^3}{3!} \vartheta'''(\theta_1 h), \quad \vartheta'''(h) = -\frac{1}{4} \cos \frac{h}{2}$$

findet man daher

$$R = -\frac{h^3}{24} \left\{ f''(\theta h) + f_{\frac{1}{2}} \cos \theta_1 \frac{h}{2} \right\}. \quad (\text{SV}) \quad (14)$$

Ganz analog erhält man im Falle des Tangentenverfahrens aus (10)

$$R = \frac{h^3}{12} \left\{ f''(\theta h) + \frac{f_0 + f_1}{2} t(\theta_1 h) \right\} \quad (\text{TV}) \quad (15)$$

mit

$$t(h) = \left( 1 + \operatorname{tg}^2 \frac{h}{2} \right) \left( 1 + 3 \operatorname{tg}^2 \frac{h}{2} \right). \quad (15a)$$

Aus (11) ergibt sich nach einigen Umformungen für das Sehen-Tangenten-Verfahren der Fehler

$$R = R_s + \frac{h^3}{36} \left( \frac{f_0 + f_1}{2} - f_{\frac{1}{2}} \right) + \frac{1}{3} f(\theta_2 h) \vartheta(h),$$

wo

$$R_s = \frac{h}{6} (f_0 + 4f_{\frac{1}{2}} + f_1) - \int_0^h f dx = \frac{1}{90} \left( \frac{h}{2} \right)^5 f^{(4)}(\theta h)$$

das Restglied der Simpsonschen Formel und

$$\vartheta(h) = 2 \operatorname{tg} \frac{h}{2} + 4 \sin \frac{h}{2} - 3h$$

ein Ausdruck von der  $O(h^5)$  ist. Ferner ist bekanntlich

$$\frac{f_0 + f_1}{2} - f_{\frac{1}{2}} = f''(\theta_1 h) \frac{h^2}{8}$$

und nach dem Taylorschen Theorem

$$\vartheta(h) = \frac{h^5}{5!} \cdot \frac{d}{dh} \vartheta(h) \Big|_{\theta_3 h}.$$

Führt man hier noch die Differentiation aus, so erhält man insgesamt

$$R = \frac{1}{90} \left( \frac{h}{2} \right)^5 \left\{ f^{(4)}(\theta h) + 10 f''(\theta_1 h) + f(\theta_2 h) t \left( \theta_3 \frac{h}{2} \right) \right\} \quad (\text{STV}) \quad (16)$$

mit

$$t(h) = 8(1 + \operatorname{tg}^2 h) \left( 1 + \frac{15}{2} \operatorname{tg}^2 h + \frac{15}{2} \operatorname{tg}^4 h \right) + \cos h. \quad (16a)$$

## Kapitel II

### Die Integrationsverfahren im Falle von Differentialgleichungen $n$ -ter Ordnung und rekursive Abschätzung des Fehlers

#### § 4. Bezeichnungen und Voraussetzungen

9. Wir legen für das Folgende die Differentialgleichung  $n$ -ter Ordnung in der nach der höchsten Ableitung aufgelösten Form

$$y^{(n)} = f(x, y, \dots, y^{(n-1)}) \quad (17)$$

zugrunde, wo  $f(x, y_0, \dots, y_{n-1})$  eine vorgegebene Funktion der  $n+1$  Variablen  $x, y_r$  ist. Durch die Anfangsbedingungen

$$y^{(v)}(x_0) = y_0^{(v)} \quad (v = 0, 1, \dots, n-1) \quad (17a)$$

sei eindeutig die Lösungsfunktion  $y(x)$  festgelegt.

Um die Schreibweise möglichst zu vereinfachen, fassen wir das System der  $n$  Variablen  $y_r$  zum Vektor  $Y = (y_0, \dots, y_{n-1})$  zusammen.

Sei wieder

$$x_r = x_0 + r h \quad (h > 0) \quad (18)$$

eine Folge von äquidistanten Argumentsstellen, wo  $r$  die Zahlenfolge  $0, 1, 2, \dots$ , gelegentlich auch die Folge  $0, 1/2, 1, 3/2, \dots$  durchläuft. (Im ersten Falle werden wir an Stelle von  $r$  den Index  $\varrho$  verwenden.) Die exakten Werte der Lösungsfunktion und ihrer Ableitungen an den Stellen (18) werden mit  $y^{(v)}(x_r)$  bezeichnet, Näherungswerte mit  $y_r^{(v)1}$ . Die entsprechenden Vektoren seien  $Y(x_r) = [y^{(v)}(x_r)]$ ,  $Y_r = (y_r^{(v)})$ . Analog bezeichnen  $G_r^{(v)} = (i^v / y_r^{(v)}) e^{ix_r}$  die Näherungspunkte der auf den  $v$ -ten Grammelschen Polaren der Lösungsfunktion  $y(x)$  gelegenen exakten Punkte  $G_v(x_r)$ .

Wir setzen fest, dass die Näherungswerte  $y_r^{(n)}$  aus der Differentialgleichung (17) mit Hilfe der Näherungen  $y_r^{(v)}$  ( $0 \leq v \leq n-1$ ) berechnet werden:

$$y_r^{(n)} = f_r = f(x_r, Y_r), \quad G_r^{(n)} = \frac{i^n}{f_r} e^{ix_r}. \quad (19)$$

Die Gesamtheit der  $n+1$  in Nr. 2 eingeführten radialen Strahlen  $s_r^{(v)}$  ( $v = 0, 1, \dots, n$ ) bilden einen sogenannten *Stern*  $S_r$ . Dieser ist Träger der Punkte  $G_r^{(v)}$  und  $G_v(x_r)$ . Wir nennen den Streckenzug  $G_r G_r' G_r'' \dots G_r^{(n)}$  das *r-te Diagramm*  $D_r$  und bezeichnen die Teilstrecken  $G_r^{(v)} G_r^{(v+1)}$  des Diagramms  $D_r$  mit  $d_r^{(v)}$  ( $v = 0, 1, \dots, n-1$ ) (vgl. Abb. 5).

10. Es sei  $\varepsilon_r^{(v)} = y_r^{(v)} - y^{(v)}(x_r)$  der Fehler, mit dem die  $v$ -te Ableitung der Lösungsfunktion  $y(x)$  nach  $r$  Integrationsschritten behaftet ist. Wir definieren „Fehlervektoren“  $E_r$  gemäss

$$E_r' = (|\varepsilon_r|, |\varepsilon_r'|, \dots, |\varepsilon_r^{(n-1)}|)^2. \quad (20)$$

Die Abschätzung der Fehler  $\varepsilon_r^{(v)}$  bzw. der Fehlervektoren  $E_r$  wird unter den folgenden *Voraussetzungen* durchgeführt werden.

1) Um eventuelle Abrundungen in den Anfangswerten  $y_0^{(v)}$  mit zu berücksichtigen, ist es zweckmässig, auch die  $y_0^{(v)}$  als Näherungen der Anfangswerte  $y^{(v)}(x_0)$  aufzufassen.

2) Der Strich beim Vektor  $E_r'$  soll andeuten, dass  $E_r$  selbst ein *Kolonnenvektor* ist.

Es sei  $B(x, Y)$  ein konvexer Bereich des  $(x, Y)$ -Raumes, der so beschaffen ist, dass sämtliche Punkte  $(x_r, Y_r)$ ,  $(x_r, Y(x_r))$  in  $B$  enthalten sind, falls  $x_r$  im Integrationsintervall  $J$  liegt. Wir nehmen dabei an, dass die angenäherte Integration bereits durchgeführt ist und die Vektoren  $Y_r$  infolgedessen bekannt sind. Insofern wird es sich also um eine *Abschätzung a posteriori*<sup>1)</sup> handeln.

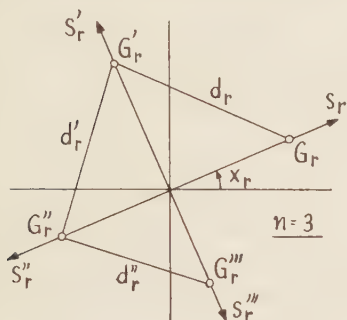


Abb. 5. Der Stern  $S_r$  als Träger der Näherungspunkte  $G_r^{(v)}$  und das Diagramm  $D_r$

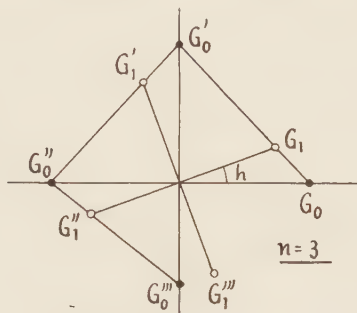


Abb. 6. Rohverfahren

(Die Abgrenzung des Bereiches  $B$  ist übrigens die einzige Stelle, an der die Kenntnis der Grössen  $y_r^{(v)}$  vorausgesetzt wird. Gelegentlich kann für  $B$  der unendliche Bereich  $x \in J, -\infty \leq Y \leq \infty$  verwendet werden, so dass in diesem Falle zum Beispiel auch eine Abschätzung a priori möglich wird.)

Wir setzen dann voraus, dass die Funktion  $f(x, Y)$  der rechten Seite der Differentialgleichung (17) in  $B$  nach allen Variablen  $q$ -mal stetig differenzierbar ist für ein  $q \geq 1$ , und sagen dann kurz, die Funktion  $f(x, Y)$  erfülle in  $B$  die Voraussetzung  $V_q$ . Es sei unter dieser Voraussetzung

$$|y^{(k)}(x)| \leq m_k \quad (x \in J, 0 \leq k \leq n + q). \quad (21)$$

Ferner möge in  $B$  gelten

$$\left| \frac{\partial f}{\partial y_v} \right| \leq k_v \quad (v = 0, 1, \dots, n-1); \quad (22)$$

dann genügt  $f$  für zwei beliebige Punkte  $(x, Y)$ ,  $(x, \bar{Y}) \in B$  einer Lipschitz-Bedingung

$$|f(x, Y) - f(x, \bar{Y})| \leq \sum_{v=0}^{n-1} k_v |y_v - \bar{y}_v|.$$

<sup>1)</sup> Die grundsätzliche Unterscheidung zwischen *Abschätzungen a posteriori* und *Abschätzungen a priori* geht auf A. OSTROWSKI (1939) zurück.

Hieraus folgt insbesondere für die Fehler  $\varepsilon_r^{(n)}$ :

$$|\varepsilon_r^{(n)}| = |f(x_r, Y_r) - f[x_r, Y(x_r)]| \leq \sum_{\nu=0}^{n-1} k_\nu |\varepsilon_r^{(\nu)}|. \quad (23)$$

Es genügt daher, die Fehler  $\varepsilon_r^{(\nu)}$  für  $0 \leq \nu \leq n-1$  abzuschätzen.

11. Die  $(n \times n)$ -Matrix

$$K = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 \\ k_0 k_1 k_2 \dots k_{n-1} \end{pmatrix}, \quad (24)$$

gebildet mit den Grössen  $k_\nu$  aus (22), wird für die Fortpflanzung des Fehlers von Bedeutung sein. Wir bezeichnen die voneinander verschiedenen *Eigenwerte* dieser Matrix mit  $\lambda_1, \dots, \lambda_s$ , ihre Vielfachheiten beziehungsweise mit  $p_1, \dots, p_s$  und denken sie so durchnummeriert, dass

$$\lambda_1 \geq |\lambda_2| \geq \dots \geq |\lambda_s| \quad (s \leq n) \quad (25)$$

gilt. Die Eigenwerte  $\lambda_\sigma$  sind bekanntlich die Wurzeln der charakteristischen Gleichung

$$\det(\lambda I - K) = \lambda^n - \sum_{\nu=0}^{n-1} k_\nu \lambda^\nu = 0. \quad (26)$$

Verschwinden nicht sämtliche  $k_\nu$ , so ist nach bekannten Sätzen der Algebra

$$\lambda_1 > 0, \quad p_1 = 1 \quad \left( \sum_{\nu} k_\nu \neq 0 \right)$$

und  $\lambda_1$  die einzige positive Wurzel. Anderenfalls ist  $\lambda_1 = 0$  eine  $n$ -fache Wurzel.

### § 5. Roh- und Sehnenverfahren

12. Die Anfangswerte (17a) der Differentialgleichung (17) legen die Anfangspunkte  $G_0^{(\nu)}$  ( $\nu = 0, 1, \dots, n-1$ ) der gesuchten Grammyschen Polaren  $G_\nu(x)$  fest und der Wert (19) von  $y_0^{(n)}$  den Punkt  $G_0^{(n)}$ . Daher kann das Diagramm  $D_0$  gezeichnet werden. Wegen (3) ist die Seite  $d_0^{(\nu)}$  dieses Diagramms Tangente an die  $\nu$ -te Polare  $G_\nu(x)$  im Punkte  $G_0^{(\nu)}$ . Um daher die Punkte  $G_\nu(x_1)$  in erster Annäherung zu bekommen, bringt man den zugehörigen Stern  $S_1$  mit dem Diagramm  $D_0$  zum Schnitt.

Genauer definieren wir das *Rohrverfahren* dahin, dass, beginnend mit  $\varrho = 0$ , der Näherungswert  $G_{\varrho+1}^{(r)}$  als Schnittpunkt des Strahls  $s_{\varrho+1}^{(p)}$  mit der Diagrammseite  $d_{\varrho}^{(v)}$  bestimmt wird:

$$G_{\varrho+1}^{(r)} = s_{\varrho+1}^{(r)} \times d_{\varrho}^{(r)} \quad (v = 0, 1, \dots, n-1; \varrho = 0, 1, 2, \dots). \quad (27)$$

Das neue Diagramm  $D_{\varrho+1}$  wird dann durch die Punkte (27) und  $G_{\varrho+1}^{(n)}$  aus (19) festgelegt (vgl. Abb. 6, die für den Fall  $\varrho = x_0 = 0$ ,  $n = 3$  gezeichnet ist).

13. Man erhält die analytische Formulierung des Rohrverfahrens aus (6), wenn man dort die Grössen  $y_{\varrho+1}$ ,  $y_{\varrho}$ ,  $f_{\varrho}$  beziehungsweise durch  $y_{\varrho+1}^{(v)}$ ,  $y_{\varrho}^{(v)}$ ,  $y_{\varrho}^{(v+1)}$  ersetzt:

$$y_{\varrho+1}^{(v)} = y_{\varrho}^{(v)} \cos h + y_{\varrho}^{(v+1)} \sin h \quad (v = 0, 1, \dots, n-1; \varrho = 0, 1, 2, \dots). \quad (\text{RV}) \quad (28)$$

Andererseits gilt für die exakte Lösungsfunktion  $y(x)$

$$y^{(v)}(x_{\varrho+1}) = y^{(v)}(x_{\varrho}) \cos h + y^{(v+1)}(x_{\varrho}) \sin h - R_v, \quad (29)$$

wo  $R_v$  der Quadraturfehler ist, den man begeht, wenn man die Funktion  $y^{(v+1)}(x)$  zwischen  $x_{\varrho}$  und  $x_{\varrho+1}$  nach (6) integriert. Aus dem Restglied (13) folgt, unter der Voraussetzung  $\Gamma_1$ , für  $R_v$  im ganzen Intervall  $J$  die Abschätzung

$$|R_v| \leq \frac{h^2}{2} q_v, \quad q_v = m_v + m_{v+2} + \frac{h}{3} m_{v+1} \quad (v = 0, 1, \dots, n-1), \quad (30)$$

in der die Konstanten  $m_k$  durch (21) definiert sind.

Subtrahiert man (29) gliedweise von (28), so erhält man für die Fehler  $\varepsilon_{\varrho}^{(v)}$  die Relationen

$$\varepsilon_{\varrho+1}^{(v)} = \varepsilon_{\varrho}^{(v)} \cos h + \varepsilon_{\varrho}^{(v+1)} \sin h + R_v \quad (v = 0, 1, \dots, n-1).$$

Daher gelten, wegen (30) und (23), die Abschätzungen

$$\left. \begin{aligned} |\varepsilon_{\varrho+1}^{(v)}| &\leq |\varepsilon_{\varrho}^{(v)}| + h |\varepsilon_{\varrho}^{(v+1)}| + \frac{h^2}{2} q_v \quad (v = 0, 1, \dots, n-2) \\ |\varepsilon_{\varrho+1}^{(n-1)}| &\leq |\varepsilon_{\varrho}^{(n-1)}| + h \sum_{v=0}^{n-1} k_v |\varepsilon_{\varrho}^{(v)}| + \frac{h^2}{2} q_{n-1} \end{aligned} \right\} \quad (31)$$

Mit Hilfe der Fehlervektoren (20) und der Matrix (24) lassen sich die Ungleichungen (31) in Matrizenform zusammenfassen zu

$$E_{\varrho+1} \leq (I + h K) E_{\varrho} + \frac{h^2}{2} Q \quad (\varrho = 0, 1, 2, \dots). \quad (\text{RV}) \quad (32)$$

Hier bedeutet  $I$  die Einheitsmatrix und  $Q$  der Vektor mit den Komponenten  $q_v$  aus (30). Im Prinzip kann man also Schranken für die Fehlervektoren  $E_{\varrho}$  rekursiv aus (32) berechnen.

14. Um das *Sehnenverfahren* auf Differentialgleichungen zu übertragen, nimmt LUDWIG (1951c) an, dass Näherungen  $y_{1/2}^{(v)}$  der Funktionswerte  $y^{(v)}(x_{1/2})$  ( $v = 0, 1, \dots, n$ ), entweder rechnerisch oder graphisch ermittelt, bereits vorliegen. Dann sind die beiden Diagramme  $D_0$  und  $D_{1/2}$  festgelegt. Gemäss dem Vorgehen in Nr. 4 bei Quadraturen wird der Stern  $S_1$  mit den *Querverbindungen*

$$q_0^{(v)} = G_{\frac{1}{2}}^{(v+1)} G_0^{(v)} \quad (v = 0, 1, \dots, n-1)$$

der beiden Diagramme  $D_0, D_{1/2}$  geschnitten. Dies liefert die Punkte  $G_1^{(v)}$  und damit das neue Diagramm  $D_1$ . Der nächste Schritt liefert die Punkte  $G_{3/2}^{(v)}$  durch den Schnitt des Sterns  $S_{3/2}$  mit den Querverbindungen der Diagramme  $D_{1/2}, D_1$  usw. Bezeichnen wir die Querverbindungen der Diagramme  $D_r, D_{r+1/2}$  mit

$$q_r^{(v)} = G_{r+\frac{1}{2}}^{(v+1)} G_r^{(v)} \quad (v = 0, 1, \dots, n-1),$$

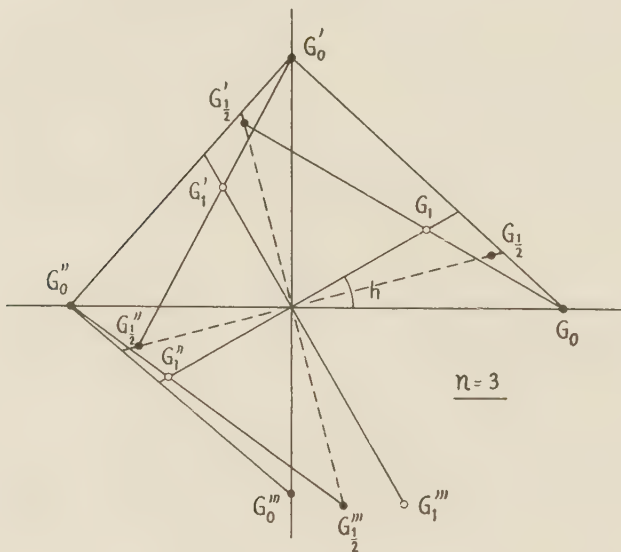


Abb. 7. Sehnenverfahren

so ist demnach allgemein<sup>1)</sup>

$$G_{r-1}^{(v)} - q_r^{(v)} \times s_{r-1}^{(v)} \quad \left( v = 0, 1, \dots, n-1; r = 0, \frac{1}{2}, 1, \frac{3}{2}, \dots \right).$$

(Vgl. Abb. 7, wo der Fall  $r = x_0 = 0, n = 3$  gezeichnet ist.)

<sup>1)</sup> Diese Konstruktion bekommt eigentlich den Charakter eines „Ganzschritt“-Verfahrens, wenn  $h$  durch  $2h$  ersetzt wird. Wir haben die Schrittweite mit  $h$  bezeichnet, damit die nachfolgenden Halbschrittverfahren mit dem Sehnenverfahren besser verglichen werden können (§ 10).

15. Das so definierte Sehnungsverfahren ist wegen (9) durch die Gleichungen

$$y_{r+1}^{(\nu)} = y_r^{(\nu)} + 2 y_{r+\frac{1}{2}}^{(\nu+1)} \sin \frac{h}{2} \quad \left( \nu = 0, 1, \dots, n-1; r = 0, \frac{1}{2}, 1, \frac{3}{2}, \dots \right) \quad (\text{SV}) \quad (33)$$

charakterisiert, während für die exakten Funktionen  $y^{(\nu)}(x)$  die Relationen

$$y^{(\nu)}(x_{r+1}) = y^{(\nu)}(x_r) + 2 y^{(\nu+1)}(x_{r+\frac{1}{2}}) \sin \frac{h}{2} - R_\nu \quad (34)$$

gelten. Unter der Voraussetzung  $V_2$  können die Grössen  $R_\nu$  als Quadraturfehler des Sehnungsverfahrens (9) nach (14) abgeschätzt werden:

$$|R_\nu| \leq \frac{h^3}{24} q_\nu, \quad q_\nu = m_{\nu+1} + m_{\nu+3} \quad (\nu = 0, 1, \dots, n-1). \quad (35)$$

Aus (33) und (34) ergibt sich analog, wie in der Nr. 13 im Falle des Rohverfahrens, die rekursive Fehlerabschätzung

$$E_{r+1} \leq h K E_{r+\frac{1}{2}} + E_r + \frac{h^3}{24} Q \quad \left( r = 0, \frac{1}{2}, 1, \frac{3}{2}, \dots \right), \quad (36)$$

wo der Vektor  $Q = (q_r)$  jetzt durch (35) definiert ist. Setzt man

$$q = 2r, \quad E_r = E_{q/2} = \bar{E}_q, \quad (37)$$

so transformiert sich (36) in die Ungleichung

$$\bar{E}_{q+2} \leq h K \bar{E}_{q+1} + \bar{E}_q + \frac{h^3}{24} Q \quad (q = 0, 1, 2, \dots). \quad (\text{SV}) \quad (38)$$

## § 6. Tangenten- und Sehnen-Tangenten-Verfahren

16. Die Quadraturformel (10) des Tangentenverfahrens legt den Gedanken nahe, im Falle von Differentialgleichungen einen *Iterationsprozess* anzusetzen. Wir bezeichnen generell den Näherungswert  $y_r^{(\nu)}$  nach dem  $j$ -ten Iterationsschritt mit  $y_{r,j}^{(\nu)}$  und den entsprechenden Näherungspunkt der  $\nu$ -ten Orthopolaren mit  $G_{r,j}^{(\nu)}$ . Für den endgültigen Näherungswert  $y_{r,\infty}^{(\nu)}$  schreiben wir, falls die Iteration konvergiert, wieder kurz  $y_r^{(\nu)}$ :

$$y_r^{(\nu)} = y_{r,\infty}^{(\nu)}. \quad (39)$$

Entsprechend zu (19) vereinbaren wir hier für alle  $r$  und  $j$ :

$$y_{r,j}^{(n)} = f_{r,j} = f(x_r, Y_{r,j}), \quad G_{r,j}^{(n)} = \frac{i^n}{f_{r,j}} e^{i x_r}. \quad (40)$$

Die Iteration, die aus den Näherungspunkten  $G_{\varrho}^{(v)}$  in der Grenze die Punkte  $G_{\varrho+1}^{(v)}$  liefern soll, wird, in Anlehnung an das Tangentenverfahren der Quadratur, folgendermassen durchgeführt. Wir wählen, um mit Näherungen von  $G_v(x_{\varrho+1})$  provisorisch arbeiten zu können, auf dem Stern  $S_{\varrho+1}$  beliebige Punkte  $G_{\varrho+1,0}^{(v)}$ , am zweckmässigsten natürlich die Punkte, die durch das Rohverfahren (27) festgelegt sind. Es mögen  $P_{\varrho+1/2}^{(v)}$  ( $v = 0, 1, \dots, n-1$ ) die Punkte bezeichnen, die durch den Schnitt des Sternes  $S_{\varrho+1/2}$  mit dem Diagramm  $D_{\varrho}$  entstehen, also die Näherungspunkte des Rohverfahrens bei halber Schrittlänge. Schneiden wir dann den Stern  $S_{\varrho+1}$  mit den Querverbindungen

$$q_{\varrho+\frac{1}{2},0}^{(v)} = G_{\varrho+1,0}^{(v+1)} P_{\varrho+\frac{1}{2}}^{(v)} \quad (v = 0, 1, \dots, n-1),$$

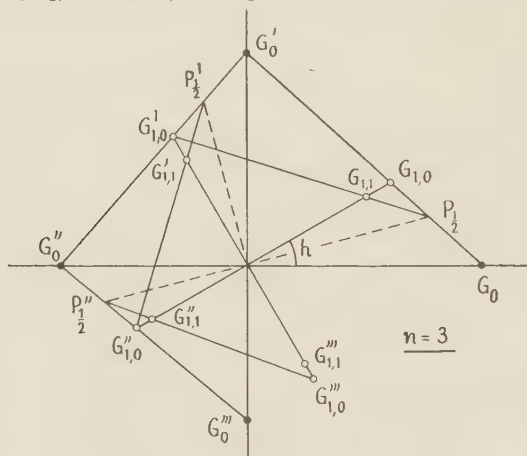


Abb. 8. Tangentenverfahren

so erhalten wir genauere Näherungspunkte  $G_{\varrho+1,1}^{(v)}$  und aus (40) den Punkt  $G_{\varrho+1,1}^{(n)}$ . Wiederholt man dieselbe Konstruktion immer wieder unter Verwendung der neuesten Näherungen, so läuft dies auf die Iteration

$$\left. \begin{aligned} G_{\varrho+1,j+1}^{(v)} &= q_{\varrho+\frac{1}{2},j}^{(v)} \times s_{\varrho+1}^{(v)} \quad (j = 0, 1, 2, \dots), \\ G_{\varrho+1}^{(v)} &= G_{\varrho+1,\infty}^{(v)} \quad (v = 0, 1, \dots, n-1; \varrho = 0, 1, 2, \dots) \end{aligned} \right\} \quad (41)$$

hinaus, wo  $q_{\varrho+\frac{1}{2},j}^{(v)}$  die Querverbindungen  $G_{\varrho+1,j}^{(v+1)} P_{\varrho+\frac{1}{2}}^{(v)}$  bedeuten<sup>1)</sup>. (Vgl. Abb. 8, in der ein einziger Iterationsschritt für  $\varrho = x_0 = 0$ ,  $n = 3$  gezeichnet ist.)

<sup>1)</sup> Die Iteration in dieser Form ist von LUDWIG (1951c) angegeben worden, während die ursprüngliche Iteration von GRAMMEL nur die beiden höchsten Ableitungen erfasst.

Aus (10) ergibt sich unmittelbar die analytische Fassung des Tangentenverfahrens (41):

$$y_{\varrho+1, j+1}^{(v)} = y_{\varrho}^{(v)} + (y_{\varrho}^{(v+1)} + y_{\varrho+1, j}^{(v+1)}) \operatorname{tg} \frac{h}{2} \\ (\nu = 0, 1, \dots, n-1; j = 0, 1, 2, \dots). \quad (\text{TV}) \quad (42)$$

Die Schrittlänge  $h$  darf freilich nicht mehr ganz willkürlich gewählt werden, wenn man sicher sein will, dass die Iteration überhaupt konvergiert.

**17. Satz.** *Unter der Bedingung*

$$\lambda_1 \operatorname{tg} \frac{h}{2} < 1 \quad (43)$$

*konvergiert das Tangentenverfahren (42) für beliebige Anfangswerte  $y_{\varrho+1, 0}^{(v)}$  der Iteration.*

*Beweis.* Mit der Bezeichnung

$$\delta_j^{(v)} = y_{\varrho-1, j+1}^{(v)} - y_{\varrho+1, j}^{(v)},$$

in der wir den Index  $\varrho+1$  der Kürze halber unterdrücken, folgen aus (42) die Relationen

$$\delta_j^{(v)} = \delta_{j-1}^{(v+1)} \operatorname{tg} \frac{h}{2} \quad (\nu = 0, 1, \dots, n-1)$$

und hieraus die Abschätzungen

$$\left. \begin{aligned} |\delta_j^{(v)}| &\leq |\delta_{j-1}^{(v+1)}| \operatorname{tg} \frac{h}{2} \quad (\nu = 0, 1, \dots, n-2), \\ |\delta_j^{(n-1)}| &= |f_{\varrho+1, j} - f_{\varrho+1, j-1}| \operatorname{tg} \frac{h}{2} \leq \sum_{\nu=0}^{n-1} k_{\nu} |\delta_{j-1}^{(\nu)}| \operatorname{tg} \frac{h}{2}. \end{aligned} \right\} \quad (44)$$

Mit Hilfe der Vektoren

$$A_j = (|\delta_j^{(v)}|) \quad (j = 0, 1, 2, \dots)$$

schreibt sich (44) in der Form  $A_j \leq (K \operatorname{tg} \frac{h}{2})^{j-1} A_1$ , woraus folgt

$$A_j \leq \left(K \operatorname{tg} \frac{h}{2}\right)^j A_0. \quad (45)$$

Für die Konvergenz der Iteration (42) ist notwendig und hinreichend, dass die Differenz  $y_{\varrho+1, j+p}^{(v)} - y_{\varrho+1, j}^{(v)}$  für beliebiges  $p$  bei hinreichend grossem  $j$  beliebig klein wird. Diese Differenz ist aber absolut höchstens gleich  $\sum_{i=j}^{j+p-1} |\delta_i^{(v)}|$ , während andererseits wegen (45)

$$\sum_{i=j}^{j+p-1} A_i \leq \sum_{i=j}^{\infty} \left(K \operatorname{tg} \frac{h}{2}\right)^i A_0 = \left(I - K \operatorname{tg} \frac{h}{2}\right)^{-1} \left(K \operatorname{tg} \frac{h}{2}\right)^j A_0 \quad (46)$$

gilt. Da der absolut grösste Eigenwert der Matrix  $K \operatorname{tg} h/2$  nach (43)  $< 1$  ist, strebt die rechte Seite in (46) mit  $j \rightarrow \infty$  gegen Null, womit die Behauptung bewiesen ist.

Ausserdem zeigt (46), dass, falls die Elementarteiler von  $K$  einfach sind,

$$y_{\varrho, j}^{(v)} - y_{\varrho}^{(v)} = O\left[\left(\lambda_1 \operatorname{tg} \frac{h}{2}\right)^j\right]$$

ist, die Iteration also im allgemeinen rasch konvergiert.

18. Für die Näherungen (39) gilt nun gemäss (42)

$$y_{\varrho+1}^{(v)} = y_{\varrho}^{(v)} + (y_{\varrho}^{(v+1)} + y_{\varrho+1}^{(v+1)}) \operatorname{tg} \frac{h}{2} \quad (v = 0, 1, \dots, n-1), \quad (47)$$

während die exakte Funktion  $y(x)$  den Relationen

$$y^{(v)}(x_{\varrho+1}) = y^{(v)}(x_{\varrho}) + [y^{(v+1)}(x_{\varrho}) + y^{(v+1)}(x_{\varrho+1})] \operatorname{tg} \frac{h}{2} - R_v \quad (48)$$

genügt. Die Quadraturfehler  $R_v$  lassen sich unter der Voraussetzung  $V_2$  nach (15) abschätzen zu

$$|R_v| \leq \frac{h^3}{12} q_v, \quad q_v = m_{v+1} t(h) + m_{v+3} \quad (v = 0, 1, \dots, n-1), \quad (49)$$

wo  $t(h)$  durch (15a) definiert ist. Aus (47) und (48) erhält man für die Fehlervektoren  $E_{\varrho}$  ähnlich wie in den Nrn. 13 und 15 die Ungleichungen

$$\left(I - K \operatorname{tg} \frac{h}{2}\right) E_{\varrho+1} \leq \left(I + K \operatorname{tg} \frac{h}{2}\right) E_{\varrho} + \frac{h^3}{12} Q \quad (\varrho = 0, 1, 2, \dots), \quad (\text{TV}) \quad (50)$$

worin für  $Q = (q_v)$  der Vektor mit den Komponenten  $q_v$  aus (49) zu nehmen ist.

19. Die Ludwigsche Kombination des Sehnenvfahrens mit dem Tangentenverfahren zum *Sehnen-Tangenten-Verfahren* wird, wegen des kleinen Quadraturfehlers (16), das theoretisch weitaus genaueste Verfahren abgeben. Auch hier wird vorausgesetzt, dass Näherungen  $y_{1/2}^{(v)}$  von  $y^{(v)}(x_{1/2})$  bereits bekannt sind. Ist dann die Konstruktion schon bis zu den Näherungen  $y_{r+1/2}^{(v)}$  fortgeschritten, so bestimmt man zunächst aus den Punkten  $G_r^{(v)}, G_{r+1/2}^{(v)}$  mit Hilfe des Sehnenvfahrens (33) die Näherungspunkte  $\hat{G}_{r+1}^{(v)}$ :

$$\hat{y}_{r+1}^{(v)} = y_r^{(v)} + 2 y_{r+\frac{1}{2}}^{(v+1)} \sin \frac{h}{2} \quad (v = 0, 1, \dots, n-1). \quad (51)$$

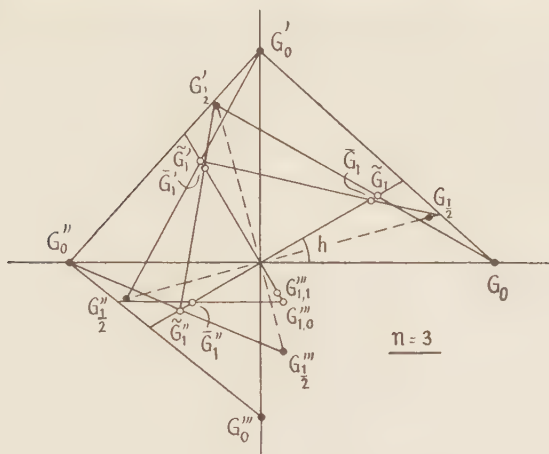


Abb. 9. Sehnentangenten-Verfahren

Es ist zweckmässig, diese Grössen als Anfangswerte einer nun anschliessenden Iteration zu benutzen, also zu setzen:

$$\tilde{y}_{r+1}^{(v)} = y_{r+1,0}^{(v)} \quad (v = 0, 1, \dots, n), \quad (52)$$

wo  $y_{r+1,0}^{(n)}$  insbesondere durch (40) gegeben ist. Ein Iterationsschritt des Tangentenverfahrens möge aus den Punkten  $G_r^{(v)}$ ,  $G_{r+1,j}^{(v)}$  die Punkte  $\bar{G}_{r+1,j+1}^{(v)}$  liefern. Dann bilde man, beginnend mit  $j = 0$ , für jedes  $j$

$$y_{r+1,j+1}^{(v)} = \frac{2}{3} \tilde{y}_{r+1}^{(v)} + \frac{1}{3} \bar{y}_{r+1,j+1}^{(v)} \quad (v = 0, 1, \dots, n-1; j = 0, 1, 2, \dots).$$

Wegen (51) und (42) erhält man daraus die Iteration

$$\left. \begin{aligned} y_{r+1,j+1}^{(v)} &= y_r^{(v)} + \frac{4}{3} y_{r+\frac{1}{2}}^{(v+1)} \sin \frac{h}{2} + \frac{1}{3} (y_r^{(v+1)} + y_{r+1,j}^{(v+1)}) \operatorname{tg} \frac{h}{2} \end{aligned} \right\} \text{ (STV)} \quad (53)$$

$$(v = 0, 1, \dots, n-1; j = 0, 1, 2, \dots)$$

mit beispielsweise (52) als Anfangswerte für  $j = 0$ .

20. Die Betrachtungen, die sich an die Relationen (42) angeschlossen haben, lassen sich in ganz analoger Weise auf (53) übertragen, so dass wir uns auf die Feststellung der entsprechenden Resultate beschränken können.

**Satz.** Unter der Bedingung

$$\frac{\lambda_1}{3} \operatorname{tg} \frac{h}{2} < 1 \quad (54)$$

konvergiert das Sehnen-Tangenten-Verfahren (53) für beliebige Anfangswerte  $y_{r+1,0}^{(v)}$  der Iteration.

Definieren wir den Vektor  $Q = (q_v)$  durch

$$\left. \begin{aligned} q_v &= [1 + 8t(h)] m_{v+1} + 10 m_{v+3} + m_{v+5} \quad (v = 0, 1, \dots, n-1), \\ t(h) &= (1 + t^2) \left( 1 + \frac{15}{2} t^2 + \frac{15}{2} t^4 \right), \quad t = \operatorname{tg} \frac{h}{2}, \end{aligned} \right\} \quad (55)$$

so gilt unter der Voraussetzung  $V_3$  für die Fehlervektoren  $E_r$  die rekursive Abschätzung

$$\left( I - \frac{1}{3} K \operatorname{tg} \frac{h}{2} \right) E_{r+1} \leq \frac{2}{3} h K E_{r+\frac{1}{2}} + \left( I + \frac{1}{3} K \operatorname{tg} \frac{h}{2} \right) E_r + \left. \begin{aligned} &+ \frac{1}{90} \left( \frac{h}{2} \right)^5 Q \quad \left( r = 0, \frac{1}{2}, 1, \frac{3}{2}, \dots \right). \end{aligned} \right\} \quad (\text{STV}) \quad (56)$$

Mit der Substitution (37) kann (56) auf ganzzahlige Indizes  $\varrho$  umgeschrieben werden.

### Kapitel III

#### Independente Fehlerabschätzung

#### im Falle von Differentialgleichungen $n$ -ter Ordnung

##### § 7. Ein System von linearen Differenzengleichungen

21. Die rekursive Abschätzung der Fehlervektoren  $E_\varrho$  für die verschiedenen Integrationsvarianten der §§ 5 und 6 liefen in allen Fällen auf eine Differenzengleichung von der allgemeinen Form

$$(a_0 I - b_0 K) E_{\varrho+m} \leq \sum_{\mu=1}^m (a_\mu I + b_\mu K) E_{\varrho+m-\mu} + R \quad (\varrho = 0, 1, 2, \dots) \quad (57)$$

Tabelle 1.

|                                        | $m$ | $a_0$ | $b_0$                                       | $a_1$ | $b_1$                           | $a_2$ | $b_2$                                       | $R$                                           |
|----------------------------------------|-----|-------|---------------------------------------------|-------|---------------------------------|-------|---------------------------------------------|-----------------------------------------------|
| Rohverfahren . . .                     | 1   | 1     | 0                                           | 1     | $h$                             | —     | —                                           | $\frac{h^2}{2} Q$                             |
| Tangentenverfahren                     | 1   | 1     | $\operatorname{tg} \frac{h}{2}$             | 1     | $\operatorname{tg} \frac{h}{2}$ | —     | —                                           | $\frac{h^3}{12} Q$                            |
| Sehnenvorfahren . .                    | 2   | 1     | 0                                           | 0     | $h$                             | 1     | 0                                           | $\frac{h^3}{24} Q$                            |
| Sehnen-Tangenten-<br>Verfahren . . . . | 2   | 1     | $\frac{1}{3} \operatorname{tg} \frac{h}{2}$ | 0     | $\frac{2}{3} h$                 | 1     | $\frac{1}{3} \operatorname{tg} \frac{h}{2}$ | $\frac{1}{90} \left( \frac{h}{2} \right)^5 Q$ |

hinaus. Dabei ist  $K$  die  $(n \times n)$ -Matrix (24), während die Zahlenwerte des Parameters  $m$ , der Konstanten  $a_\tau$ ,  $b_\tau$  und die Vektoren  $R$  in Tabelle 1 zusammengestellt sind<sup>1)</sup>.

Die Komponenten  $q_\nu$  der Vektoren  $Q$  sind für die verschiedenen Integrationsvarianten verschieden zu berechnen, nämlich beziehungsweise nach den Formeln (30), (49), (35), (55). Man entnimmt Tabelle 1 die Relationen

$$a_0 = \sum_{\mu=1}^m a_\mu \neq 0, \quad b \equiv \sum_{\tau=0}^m b_\tau \neq 0, \quad a_\tau, b_\tau \geq 0 \quad (\tau = 0, 1, \dots, m). \quad (58)$$

Es soll in diesem Kapitel die Ungleichung (57) unter den allgemeinen<sup>2)</sup> Bedingungen (58) diskutiert und insbesondere für Vektoren  $E_\varrho$ , die (57) und gewissen Anfangsbedingungen genügen, Schranken verschiedener Genauigkeitsstufe hergeleitet werden.

Wir ziehen zu diesem Zweck die (57) entsprechende Vektordifferenzengleichung

$$(a_0 I - b_0 K) U_{\varrho+m} = \sum_{\mu=1}^m (a_\mu I + b_\mu K) U_{\varrho+m-\mu} + R \quad (\varrho = 0, 1, 2, \dots) \quad (59)$$

für die Vektoren  $U_\varrho$  heran. Es wird sich nämlich leicht zeigen lassen, dass geeignete Lösungsvektoren dieser Gleichung Schranken für die Vektoren  $E_\varrho$  abgeben. Daher soll in diesem Paragraphen zunächst die allgemeine Lösung der Differenzengleichung (59) bereitgestellt werden<sup>3)</sup>.

22. Wir führen das Polynom  $\varphi(\lambda, t)$  vom Grade  $m$  in  $t$  ein,

$$\varphi(\lambda, t) = (a_0 - b_0 \lambda) t^m - \sum_{\mu=1}^m (a_\mu + b_\mu \lambda) t^{m-\mu}, \quad (60)$$

gewissermassen das *charakteristische Polynom* der Gleichung (59), und setzen neben (58) voraus, dass für sämtliche Eigenwerte  $\lambda_\sigma$  der Matrix  $K$ ,  $\lambda_1 \geq |\lambda_2| \geq \dots \geq |\lambda_s|$ , die Diskriminante  $D(\lambda)$  dieses Polynoms sowie Anfangs- und Endkoeffizienten nicht verschwinden. Genauer setzen wir voraus, dass

$$D(\lambda_\sigma) \neq 0, \quad a_\tau - b_\tau \lambda_1 > 0 \quad (\tau = 0, m)^4) \quad (61)$$

<sup>1)</sup> Die Bezeichnungen Tangenten- bzw. Seilungsverfahren in Tabelle 1 beziehen sich auf das Grammelsche Verfahren (vgl. die Bemerkung am Schlusse der Nr. 7).

<sup>2)</sup> Die Bedingungen (58) sind auch bei gewissen *numerischen* Integrationsverfahren, sofern sie auf der Zerlegung der Differentialgleichung in ein System erster Ordnung beruhen, erfüllt.

<sup>3)</sup> Für die zugehörigen Beweise vgl. GAUTSCHI (1954).

<sup>4)</sup> Für  $m = 1$  entfällt die erste dieser Bedingungen. Die Bedingungen (61) sind für die Werte in Tabelle 1 stets erfüllt, wenn  $\lambda_1 \operatorname{tg} h/2 < 1$ , also die Konvergenzbedingung (43) vorausgesetzt wird.

gilt. Aus den beiden Ungleichungen in (61) folgt dann, für  $\tau = 0$  und  $\tau = m$ ,

$$|\lambda_\sigma| b_\tau \leq \lambda_1 b_\tau < a_\tau, \quad \text{also} \quad (a_0 - b_0 \lambda_\sigma) (a_m + b_m \lambda_\sigma) \neq 0 \quad (\sigma = 1, \dots, s).$$

Daher haben die Gleichungen  $\varphi(\lambda_\sigma, t) = 0$  je  $m$  Wurzeln  $t_{\sigma, \mu}$  ( $\mu = 1, \dots, m$ ), die alle *einfach* sind und *nicht verschwinden*. Hievon seien die Wurzeln  $t_{1, \mu}$  nach fallenden Beträgen geordnet,

$$t_{1,1} \geq |t_{1,2}| \geq \dots \geq |t_{1,m}|, \quad (62)$$

die übrigen  $t_{\sigma, \mu}$  ( $\sigma > 1$ ) hingegen in beliebiger Ordnung.

Wir definieren die Funktionen  $v_{\sigma, i}^\mu(\varrho)$  durch

$$v_{\sigma, i}^\mu = \frac{1}{i!} \cdot \frac{d^i}{d\lambda_\sigma^i} v_{\sigma, 0}^\mu \quad (i = 1, 2, \dots), \quad v_{\sigma, 0}^\mu = t_{\sigma, \mu}^\varrho. \quad (63)$$

Die Ableitungen von  $t_{\sigma, \mu}$  nach  $\lambda_\sigma$  sind dabei durch totale Differentiation der impliziten Gleichung  $\varphi(\lambda_\sigma, t) = 0$  nach  $\lambda_\sigma$  zu bestimmen. Die ersten Funktionen in (63) lauten dann

$$\begin{aligned} v_{\sigma, 0}^\mu(\varrho) &= t_{\sigma, \mu}^\varrho, \\ v_{\sigma, 1}^\mu(\varrho) &= \varrho t_{\sigma, \mu}^{\varrho-1} t'_{\sigma, \mu}, \\ v_{\sigma, 2}^\mu(\varrho) &= \binom{\varrho}{2} t_{\sigma, \mu}^{\varrho-2} t'^2_{\sigma, \mu} + \frac{1}{2} \varrho t_{\sigma, \mu}^{\varrho-1} t''_{\sigma, \mu}, \\ &\dots \end{aligned}$$

Ähnlich seien die Vektoren  $C_{\sigma, l}$  gemäss

$$C_{\sigma, l} = \frac{1}{(l-1)!} \cdot \frac{d^{l-1}}{d\lambda_\sigma^{l-1}} C_{\sigma, 1} \quad (l = 1, 2, \dots), \quad C'_{\sigma, 1} = (1, \lambda_\sigma, \dots, \lambda_\sigma^{n-1})$$

definiert.

Dann ist

$$U_h(\varrho) = \sum_{\sigma=1}^s \sum_{l=1}^{p_\sigma} w_{\sigma l}(\varrho) C_{\sigma, l}, \quad w_{\sigma l}(\varrho) = \sum_{\mu=1}^m \sum_{i=0}^{p_\sigma-l} c_{\sigma, i+l}^\mu v_{\sigma, i}^\mu(\varrho), \quad (64)$$

wo  $c_{\sigma, j}^\mu$  ( $\sigma = 1, \dots, s$ ;  $j = 1, \dots, p_\sigma$ ;  $\mu = 1, \dots, m$ )  $mn$  beliebige Konstanten sind, die *allgemeine Lösung* der (59) entsprechenden *homogenen* Gleichung.

Für den häufig eintretenden Fall, in dem die Matrix  $K$  lauter *verschiedene* Eigenwerte  $\lambda_\sigma$  besitzt ( $p_\sigma = 1$ ), ist  $U_h(\varrho)$  einfach gleich

$$U_h(\varrho) = \sum_{\sigma=1}^s \left( \sum_{\mu=1}^m c_{\sigma, \mu} t_{\sigma, \mu}^\varrho \right) C_{\sigma, 1} \quad (p_\sigma = 1)$$

mit beliebigen Konstanten  $c_{\sigma, \mu}$ .

23. Eine Partikulärlösung der *inhomogenen* Gleichung (59) sieht verschieden aus, je nachdem ob  $K$  einen Eigenwert 0 hat oder nicht. Ist  $\lambda_s \neq 0$ , so ist, wegen der ersten Relation in (58), der konstante Vektor  $U_p(\varrho) = -K^{-1} R/b$  eine Partikulärlösung. Ist hingegen  $\lambda_s = 0$ , so kommen noch polynomiale Bestandteile hinzu.

Sei allgemein

$$p = \begin{cases} 0 & (\lambda_s \neq 0) \\ p_s & (\lambda_s = 0) \end{cases}$$

die „Vielfachheit“ des Eigenwertes 0. Dann gilt, wegen (26),

$$k_0 = \dots = k_{p-1} = 0, \, k_p \neq 0.$$

Wir bezeichnen mit  $K_{n-p}$  die  $K$  nachgebildete  $(n-p)$ -reihige Matrix, deren letzte Zeile mit  $k_p$  beginnt:

$$K_{n-p} = \begin{pmatrix} 0 & 1 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \\ k_p & k_{p+1} \dots & k_{n-1} \end{pmatrix}, \tag{65 a}$$

$$K_{n-p}^{-1} = \begin{pmatrix} -k_{p+1}/k_p & -k_{p+2}/k_p & \dots & -k_{n-1}/k_p & 1/k_p \\ 1 & 0 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 1 & 0 \end{pmatrix}. \tag{65 b}$$

Wegen den ersten Relationen in (58) und (61) hat das Polynom

$$\varphi(0, t) = a_0 t^m - \sum_{\mu=1}^m a_\mu t^{m-\mu}$$

im Falle  $p > 0$  eine einfache Wurzel  $t = 1$ :

$$\varphi(0, t) = (t - 1) \varphi^*(t), \, \varphi^*(1) \neq 0.$$

Sei  $1/\varphi^*(1+t) = \gamma_0 + \gamma_1 t + \gamma_2 t^2 + \dots$  ( $\gamma_0 \neq 0$ ). Mit  $E$  und  $\Delta$  mögen, wie üblich, der Verschiebungs- bzw. Differenzenoperator bezeichnet werden und mit  $B$  der Operator  $B = \sum_{\tau=0}^m b_\tau E^{m-\tau}$ . Ferner sei  $\Delta^{-1}$  der Operator, der angewandt auf ein Polynom  $\binom{q}{r}$  das Polynom  $\binom{q}{r+1}$  liefert. Dann wird durch die rekursive Beziehung

$$\omega_i(q) = B \Delta^{-1} \sum_{r=0}^i \gamma_r A^r \omega_{i-1}(q) \quad (i = 1, 2, \dots), \quad \omega_0 = \gamma_0 q \tag{66}$$

eindeutig eine Folge von Polynomen  $\omega_i(q)$  definiert:

$$\begin{aligned} \omega_0(q) &= \gamma_0 q, \\ \omega_1(q) &= \gamma_0 B \left\{ \gamma_0 \binom{q}{2} + \gamma_1 q \right\}, \\ \omega_2(q) &= \gamma_0 B^2 \left\{ \gamma_0^2 \binom{q}{3} + 2 \gamma_0 \gamma_1 \binom{q}{2} + (\gamma_1^2 + \gamma_0 \gamma_2) q \right\}, \\ &\dots \end{aligned}$$

Für  $m = 1$  insbesondere ist

$$\gamma_0 = -\frac{1}{a_0}, \quad \gamma_r = 0 \quad (r > 0)$$

und für  $m = 2$ , wie man leicht verifiziert,

$$\gamma_v = (-1)^{v+1} \frac{a_0^v}{(2a_0 + a_1)^{v+1}} \quad (v \geq 0).$$

Bildet man mit den Polynomen (66) die Matrix

$$\Omega(\varrho) = \left( \begin{array}{cccc|cccc} \omega_0 & \omega_1 & \dots & \omega_{p-1} & -\omega_{p-1} & 0 & \dots & 0 \\ 0 & \omega_0 & \dots & \omega_{p-2} & -\omega_{p-2} & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \omega_0 & -\omega_0 & 0 & \dots & 0 \\ \hline & & & 0 & & -\frac{1}{b} I_{n-p} & & \end{array} \right),$$

wo allgemein  $I_k$  die  $k$ -reihige Einheitsmatrix bedeutet, so ist

$$U_p(\varrho) = P(\varrho)R, \quad P(\varrho) = \Omega(\varrho) \begin{pmatrix} I_p & 0 \\ 0 & K_{n-p}^{-1} \end{pmatrix} \quad (67)$$

eine *Partikulärlösung* der *inhomogenen* Gleichung (59).

Man sieht, dass für  $p = 0$  die bereits erwähnte konstante Partikulärlösung herauskommt.

Zusammenfassend gilt der

**Hilfssatz I.** Die allgemeine Lösung  $U(\varrho)$  der Differenzengleichung (59) kann in der Form

$$U(\varrho) = U_h(\varrho) + U_p(\varrho) \quad (68)$$

geschrieben werden, wo die Vektoren  $U_h(\varrho)$  und  $U_p(\varrho)$  durch (64) bzw. (67) gegeben sind.

## § 8. Das Prinzip der Fehlerabschätzung für allgemeines $n$

24. Damit aus (57) überhaupt sukzessive Schranken für  $E_\varrho$  berechnet werden können, muss  $(a_0 I - b_0 K)^{-1} \geq 0$ <sup>1)</sup> sein. Dies ist aber unter der Voraussetzung  $\lambda_1 b_0 < a_0$  in (61) erfüllt. Denn ist  $b_0 = 0$ , so ist die Behauptung trivial. Für ein  $b_0 > 0$  folgt aber  $|\lambda_\sigma| < a_0/b_0$  für sämtliche Eigenwerte  $\lambda_\sigma$  von  $K$ . Daher ergibt sich die Behauptung aus der Liouvilleschen Entwicklung

$$(a_0 I - b_0 K)^{-1} = \frac{1}{a_0} \sum_{v=0}^{\infty} \left( \frac{b_0}{a_0} K \right)^v \geq 0$$

25. Einen ersten Überblick über das Anwachsen der  $E_\varrho$  gewinnt man durch Abschätzen einer *Norm* dieser Vektoren [RICH-

<sup>1)</sup> Für irgendeine Matrix  $A$  soll  $A \geq 0$  bedeuten, dass sämtliche Elemente dieser Matrix  $\geq 0$  sind.

<sup>2)</sup> Die bewiesene Tatsache folgt auch aus allgemeinen Sätzen von A. OSTROWSKI (1937) über sogenannte  $M$ -Matrizen.

TER (1952)]. Für einen Vektor  $X' = (x_1, \dots, x_n)$  kann hier mit Vorteil eine der „ $q$ -Normen“

$$|X|_q = \left( \sum_{v=1}^n |x_v|^q \right)^{1/q} \quad (q \geq 1)^1$$

verwendet werden. Wählen wir beispielsweise  $q = \infty$ ,  $|X|_\infty = \max_v |x_v|$ , und bezeichnet  $k$  die entsprechende Norm des Vektors  $(k_0, \dots, k_{n-1})$ , so gilt offenbar

$$|KX|_\infty \leq \kappa |X|_\infty, \quad \kappa = \begin{cases} k_0 & (n=1), \\ \max(nk, 1) & (n>1). \end{cases}$$

Aus (57) folgt aber

$$a_0 E_{\varrho+m} \leq b_0 K E_{\varrho+m} + \sum_{\mu=1}^m (a_\mu I + b_\mu K) E_{\varrho+m-\mu} + R$$

und daher für die Normen  $e_\varrho = |E_\varrho|_\infty$ ,  $r = |R|_\infty$

$$(a_0 - \kappa b_0) e_{\varrho+m} \leq \sum_{\mu=1}^m (a_\mu + \kappa b_\mu) e_{\varrho+m-\mu} + r. \quad (69)$$

In (69) hat man somit eine rekursive Abschätzung der Normen  $e_\varrho$ , vorausgesetzt, dass  $\kappa b_0 < a_0$  ist. Die  $e_\varrho$  selbst können dann in bekannter Weise durch ein Potenzgesetz abgeschätzt werden<sup>2)</sup>.

26. Genaueren Aufschluss über die *einzelnen Komponenten* von  $E_\varrho$  erhält man auf dem Wege über die Differenzengleichung (59). Es gilt nämlich der

**Satz I.** *Ist der Vektor  $U(\varrho)$  eine Lösung der Differenzengleichung (59) und ist für  $\varrho = 0, 1, \dots, m-1$*

$$U(\varrho) \geq E_\varrho, \quad (70)$$

*so gelten die Ungleichungen (70) auch für sämtliche  $\varrho \geq m$ .*

Der Beweis ergibt sich durch vollständige Induktion. Denn (70) ist wahr für  $\varrho = 0, 1, \dots, m-1$ . Sei die Behauptung richtig für  $0 \leq \varrho \leq \varrho_0$ ,  $\varrho_0 \geq m-1$ . Setzt man dann  $V(\varrho) = E_\varrho - U(\varrho)$ , so ergeben (57) und (59) für  $\varrho = \varrho_0 + 1 - m$

$$(a_0 I - b_0 K) V(\varrho_0 + 1) \leq \sum_{\mu=1}^m (a_\mu I + b_\mu K) V(\varrho_0 + 1 - \mu) \leq 0, \\ V(\varrho_0 + 1) \leq 0,$$

also die Behauptung für  $\varrho_0 + 1$ . Damit ist Satz I bewiesen.

<sup>1)</sup> Die Bezeichnung  $|X|_q$  für die  $q$ -Norm einer Matrix  $X$  geht auf A. OSTROWSKI zurück.

<sup>2)</sup> Vgl. VON MISES (1930) oder auch COLLATZ (1951).

27. Je nach der Art, wie man die Ungleichungen (70) für  $\varrho = 0, 1, \dots, m-1$  befriedigt, ergeben sich nach Satz I verschieden genaue Schranken  $U(\varrho)$  für die Fehlervektoren  $E_\varrho$ . Recht genaue Schranken erhält man, wenn,

$$\varepsilon = \max_{0 \leq \varrho \leq m-1} e_\varrho, \quad e_\varrho = |E_\varrho|_\infty \quad (71)$$

gesetzt, die allgemeine Lösung  $U(\varrho)$  in (68) so spezialisiert wird, dass

$$U(\varrho) = \varepsilon \quad (\varrho = 0, 1, \dots, m-1) \quad (72)$$

gilt. (72) stellt ein lineares Gleichungssystem  $mn$ -ter Ordnung dar für die  $mn$  freien Konstanten  $c_{\sigma,j}^\mu$  in (64).

Freilich wäre es zu umständlich, die Auflösung des Gleichungssystems (72) in dieser allgemeinen Form durchzuführen. Im gegebenen Einzelfall jedoch, besonders wenn  $m$  nur gleich 1 oder 2 und  $n$  nicht zu gross ist, führt dieser Weg mit erträglichem Aufwand an Rechnung zum Ziel. Für die praktisch wichtigen Fälle  $n=1$ ,  $n=2$  und für  $m=1,2$  werden in § 9 explizite Schranken, die sich so ergeben, zusammengestellt.

28. Der folgenden, zwar weniger genauen, aber für praktische Zwecke oft bequemeren Abschätzung liegt folgender Hilfssatz zugrunde:

**Hilfssatz II.** Bezeichnet  $t_{1,1}$  die unter den Annahmen (58), (61) und (62) positive Wurzel des Polynoms  $\varphi(\lambda_1, t)$  und ist  $t_\lambda$  für eine beliebige komplexe Zahl  $\lambda$  mit  $|\lambda| \leq \lambda_1$  eine Wurzel von  $\varphi(\lambda, t) = 0$ , so gilt

$$|t_\lambda| \leq t_{1,1} \quad (|\lambda| \leq \lambda_1). \quad (73)$$

*Beweis.* Aus

$$(a_0 - b_0 \lambda) t_\lambda^m = \sum_{\mu=1}^m (a_\mu + b_\mu \lambda) t_\lambda^{m-\mu}$$

ergibt sich wegen  $|\lambda| \leq \lambda_1$

$$(a_0 - b_0 \lambda_1) |t_\lambda|^m \leq |(a_0 - b_0 \lambda) t_\lambda^m| \leq \sum_{\mu=1}^m (a_\mu + b_\mu \lambda_1) |t_\lambda|^{m-\mu},$$

also  $\varphi(\lambda_1, |t_\lambda|) \leq 0$ . Da offenbar  $\varphi(\lambda_1, t) \leq 0$  für  $0 \leq t \leq t_{1,1}$  und  $> 0$  für  $t > t_{1,1}$  ist, folgt daraus sofort die Behauptung (73).

29. Wir setzen für das Folgende voraus, dass

$$\lambda_1 > 0 \quad (74)$$

ist, das heisst, dass nicht sämtliche Eigenwerte  $\lambda_\sigma$  von  $K$  und somit nicht sämtliche Konstanten  $k_\nu$  dieser Matrix verschwinden. Dann folgt aus dem Hilfssatz II, dass  $t_{1,1} > 1$  ist. Denn für  $\lambda = 0$  kann wegen der ersten Relation in (58)  $t_\lambda = 1$  gesetzt werden. Andererseits ist sicher  $t_{1,1} \neq 1$ , denn sonst müsste

$$\varphi(\lambda_1, 1) = -b\lambda_1 = 0$$

sein, entgegen den Annahmen  $b \neq 0$ ,  $\lambda_1 > 0$ .

Für die Wurzeln  $t_{\sigma, \mu}$  der Polynome  $\varphi(\lambda_\sigma, t)$  gilt somit

$$|t_{\sigma, \mu}| \leq t_{1,1} \quad (\sigma = 1, \dots, s; \mu = 1, \dots, m), \quad t_{1,1} > 1. \quad (75)$$

Beachtet man, dass der Bestandteil  $U_h(\varrho)$  der allgemeinen Lösung  $U(\varrho)$  in (68) sich, abgesehen von eventuellen polynomialen Faktoren, linear zusammensetzt aus den Potenzen  $t_{\sigma, \mu}^\varrho$  der Wurzeln (75) und auch  $U_p(\varrho)$  höchstens polynomiale Elemente hat, so sieht man, dass  $t_{1,1}$  allein dafür kennzeichnend ist, wie stark  $U(\varrho)$  und somit auch die Vektoren  $E_\varrho$  anwachsen, wenn  $\varrho \rightarrow \infty$ . Daher liegt es nahe, an Stelle der vollständigen Lösung  $U(\varrho)$  in Satz I eine spezielle Lösung  $U^*(\varrho)$  zu verwenden, die nur die Wurzel  $t_{1,1}$  enthält. Eine solche ergibt sich, wenn man in (64)

$$c_{1,1}^1 = c, \quad c_{\sigma, j}^\mu = 0 \quad (\sigma > 1; j = 1, \dots, p_\sigma; \mu = 1, \dots, m)$$

setzt, zu

$$U^*(\varrho) = c t_{1,1}^\varrho C_1 + P(\varrho) R.$$

Dabei ist  $C_1$  der Vektor

$$C_1' = C_{1,1}' = (1, \lambda_1, \dots, \lambda_1^{n-1}). \quad (76)$$

Die Abschätzung vollzieht sich daher in diesem Fall nachfolgendem

**Satz II.** Sei  $\lambda_1 > 0$ ,  $t_{1,1}$  die positive Wurzel von  $\varphi(\lambda_1, t) = 0$  und  $C_1$  der Vektor (76). Wird die Konstante  $c$  so bestimmt, dass für  $\varrho = 0, 1, \dots, m-1$

$$c t_{1,1}^\varrho C_1 + P(\varrho) R \geq E_\varrho \quad (77)$$

gilt, so bleiben diese Ungleichungen bestehen für sämtliche  $\varrho \geq m$ .

Im Falle  $p = 0$  läuft (77) wegen (67) auf die Ungleichungen

$$c t_{1,1}^\varrho C_1 \geq \frac{1}{b} K^{-1} R + E_\varrho \quad (77a)$$

hinaus. Nun gilt aber für das Produkt der Matrix (65b) mit dem Vektor  $R = (R_0, \dots, R_{n-1})$  offenbar

$$(K^{-1} R)' \leq \left( \frac{1}{k_0} R_{n-1}, R_0, \dots, R_{n-2} \right).$$

Daher kann in diesem Falle

$$c = \left\{ \begin{array}{ll} r^* + \varepsilon & (\lambda_1 \geq 1) \\ \lambda_1^{1-n}(r^* + \varepsilon) & (\lambda_1 < 1) \end{array} \right\} \quad (p=0) \quad (78)$$

$$r^* = \frac{|R|_\infty}{b k_0} \gamma, \quad \gamma = \left\{ \begin{array}{ll} 1 & (n=1) \\ \text{Max}(k_0, 1) & (n>1) \end{array} \right\}$$

gesetzt werden, wo  $\varepsilon$  das Maximum in (71) bedeutet.

### § 9. Explizite Fehlerschranken für $n=1$ und $n=2$

30. Um unnötige Komplikationen zu vermeiden, passen wir die allgemeine Differenzenungleichung (57) etwas mehr an die in Tabelle 1 gegebenen Verhältnisse an, indem wir bemerken, dass in allen Fällen  $a_0 = a_m = 1$  und für  $m=2$  noch  $a_1 = 0$ ,  $b_2 = b_0$  ist. Ausserdem gelten die Ungleichungen (57), je nachdem, ob  $m=1$  oder  $m=2$  ist, für die durch (20) definierten Fehlervektoren  $E_\varrho$  oder für die gemäss (37) transformierten Vektoren  $\bar{E}_\varrho = E_{\varrho/2}$  ( $\varrho = 0, 1, 2, \dots$ ).

Wir beschränken uns daher darauf, die Fehlervektoren  $E_\varrho$  für  $n=1$  und  $n=2$  abzuschätzen, wenn sie entweder einer Differenzenungleichung

$$(I - b_0 K) E_{\varrho+1} \leq (I + b_1 K) E_\varrho + R \quad (\varrho = 0, 1, 2, \dots) \quad (m=1) \quad (79a)$$

oder einer Ungleichung von der Form

$$\left. \begin{array}{l} (I - b_0 K) \bar{E}_{\varrho+2} \leq b_1 K \bar{E}_{\varrho+1} + (I + b_0 K) \bar{E}_\varrho + R, \\ \bar{E}_\varrho = E_{\frac{\varrho}{2}} \quad (\varrho = 0, 1, 2, \dots) \end{array} \right\} \quad (m=2) \quad (79b)$$

genügen. Ist im letzteren Fall  $U(\varrho)$  eine die Vektoren  $\bar{E}_\varrho$  majorisierende Lösung der zugehörigen Differenzengleichung, so gilt offenbar

$$E_\varrho \leq \bar{U}(2\varrho) \quad (\varrho = 0, 1, 2, \dots) \quad (m=2). \quad (80)$$

Für die Wurzeln  $t_{\sigma, \mu}$  der den Ungleichungen (79a) und (79b) zugeordneten charakteristischen Polynome  $\varphi(\lambda_\sigma, t)$  erhält man

$$t_{\sigma, 1} = \frac{1 + b_1 \lambda_\sigma}{1 - b_0 \lambda_\sigma} \quad (m=1),$$

$$t_{\sigma, \mu} = \frac{\frac{b_1}{2} \lambda_\sigma + (-1)^{\mu-1} \sqrt{1 + \left[ \left( \frac{b_1}{2} \right)^2 - b_0^2 \right] \lambda_\sigma^2}}{1 - b_0 \lambda_\sigma} \quad (m=2; \mu=1, 2). \quad (81)$$

31. Wir betrachten zunächst den Fall  $n = 1$ . Hier reduziert sich die Matrix  $K$  auf die skalare Grösse  $k_0$ , so dass nur ein einziger „Eigenwert“

$$\lambda_1 = k_0$$

existiert. Offenbar kann  $k_0 > 0$  angenommen werden, da man für  $k_0 = 0$  auf den Fall der Quadratur zurückkommt.

Zur Abschätzung des Fehlers  $E_\varrho - \varepsilon_\varrho$  reicht in diesem einfachen Fall der Satz II vollständig aus. Mit der Bedeutung (71) von  $\varepsilon$  kann nach (78)  $c = (R_0/bk_0) + \varepsilon$  gesetzt werden, wo  $b$  die in (58) definierte Konstante und  $R_0$  die einzige Komponente von  $R$  ist. Unter Beachtung von (80) erhalten wir daher für  $m = 1, 2$  aus (77 a) die Fehlerabschätzung

$$|\varepsilon_\varrho| \leq \frac{R_0}{b k_0} (t_{1,1}^{m\varrho} - 1) + \varepsilon t_{1,1}^{m\varrho} \quad (m = 1, 2; \varrho = 0, 1, 2, \dots) \quad (n = 1), \quad (82)$$

in der  $t_{1,1}$  eine der Wurzeln in (81) ist.

32. Im Falle  $n = 2$  berechnen sich die Eigenwerte  $\lambda_\sigma$  aus der Gleichung

$$\lambda^2 - k_1 \lambda - k_0 = 0 \quad (k_0, k_1 \geq 0). \quad (83)$$

Falls  $k_0 + k_1 \neq 0$  ist, existieren zwei voneinander verschiedene Lösungen  $\lambda_1, \lambda_2, |\lambda_2| \leq \lambda_1, \lambda_2 \leq 0$ . Sonst hat (83) die doppelte Wurzel  $\lambda_1 = 0$ . Wir definieren mit Hilfe dieser Eigenwerte und der Wurzeln  $t_{\sigma,\mu}$  aus (81) die Ausdrücke

$$T_\sigma(\varrho) = \begin{cases} t_{\sigma,1}^\varrho & (m = 1) \\ \frac{1}{t_{\sigma,1} - t_{\sigma,2}} [(1 - t_{\sigma,2}) t_{\sigma,1}^\varrho + (t_{\sigma,1} - 1) t_{\sigma,2}^\varrho] & (m = 2). \end{cases}$$

Sie haben als lineare Verbindungen der Potenzen  $t_{\sigma,\mu}^\varrho$  die charakteristische Eigenschaft  $T_\sigma(0) - T_\sigma(1) = 1$  für  $m = 2$  und  $T_\sigma(0) = 1$  für  $m = 1$ .

Das allgemeine Abschätzungsprinzip des § 8 ergibt, je nachdem ob die Vielfachheit  $p$  des Eigenwertes 0 gleich 0, 1 oder 2 ist, für die Fehlervektoren

$$E_\varrho = \begin{pmatrix} |\varepsilon_\varrho| \\ |\varepsilon'_\varrho| \end{pmatrix} \quad (\varrho = 0, 1, 2, \dots)$$

verschieden geartete Schranken:

$$\begin{aligned} \text{a) Es sei } p = 0, \text{ also } \lambda_1 > 0, \lambda_2 \neq 0. \text{ Dann ist, für } m = 1, 2, \\ |\varepsilon_\varrho| &\leq A T_1(m\varrho) + B T_2(m\varrho) + C + \varepsilon (A' T_1(m\varrho) + B' T_2(m\varrho)), \\ |\varepsilon'_\varrho| &\leq \lambda_1 A T_1(m\varrho) + \lambda_2 B T_2(m\varrho) + D + \varepsilon (\lambda_1 A' T_1(m\varrho) + \lambda_2 B' T_2(m\varrho)) \end{aligned} \quad (84a)$$

mit

$$A = \frac{1}{b} \cdot \frac{R_1 - \lambda_2 R_0}{\lambda_1 (\lambda_1 - \lambda_2)}, \quad B = \frac{1}{b} \cdot \frac{\lambda_1 R_0 - R_1}{\lambda_2 (\lambda_1 - \lambda_2)}, \quad C = \frac{1}{b} \cdot \frac{R_1 - (\lambda_1 + \lambda_2) R_0}{\lambda_1 \lambda_2},$$

$$D = -\frac{1}{b} R_0,$$

$$A' = \frac{1 - \lambda_2}{\lambda_1 - \lambda_2}, \quad B' = \frac{\lambda_1 - 1}{\lambda_1 - \lambda_2}.$$

b) Es sei  $p=1$ , also  $\lambda_1 > 0$ ,  $\lambda_2 = 0$ . Dann gilt

$$|\varepsilon_\varrho| \leq \frac{R_1}{b \lambda_1^2} (T_1(m\varrho) - 1) + \frac{\lambda_1 R_0 - R_1}{\lambda_1} \varrho + \frac{\varepsilon}{\lambda_1} (T_1(m\varrho) - 1) + \varepsilon,$$

$$|\varepsilon'_\varrho| \leq \frac{R_1}{b \lambda_1} (T_1(m\varrho) - 1) + \varepsilon T_1(m\varrho)^1). \quad (84b)$$

c) Schliesslich sei  $p=2$ , das heisst  $\lambda_1 = 0$ . Dann gilt

$$|\varepsilon_\varrho| \leq \frac{b R_1}{2} \varrho^2 + \left[ R_0 - \frac{R_1}{2} ((m-2)b_0 + b_1) \right] \varrho + \varepsilon (b\varrho + 1), \quad (84c)$$

$$|\varepsilon'_\varrho| \leq R_1 \varrho + \varepsilon.$$

In der Abschätzung a) sind die Konstanten  $k_0, k_1$  durch die offenbar kennzeichnenderen Konstanten  $\lambda_1, \lambda_2$  ausgedrückt worden, während in b) natürlich  $\lambda_1 = k_1$  gilt. Die Fehlerabschätzung

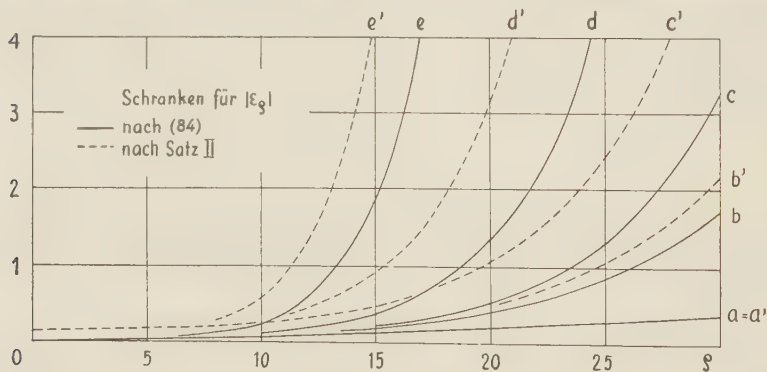


Abb. 10. a)  $\lambda_1 - \lambda_2 = 0$ , b)  $\lambda_1 = 3/2, \lambda_2 = -1$ , c)  $\lambda_1 = 2, \lambda_2 = -1/5$ ,  
d)  $\lambda_1 = 3, \lambda_2 = -1/2$ , e)  $\lambda_1 = 5, \lambda_2 = -2$

unter c) ist insofern von Interesse, als sie den Fall  $k_0 = k_1 = 0$ , also den Fall zweimaliger Quadratur, wiedergibt. Man sieht, dass in diesem Falle der Fehler nur mit der zweiten Potenz von  $\varrho$  anwächst, während er unter a) und b) exponentiell ansteigt.

<sup>1)</sup> In (84b) kann an Stelle von  $T_1(m\varrho)$  auch einfacher  $t_{1,1}^{m\varrho}$  gesetzt werden, da  $t_{1,1}^{m\varrho} \geq T_1(m\varrho)$  ist.

(84b) und (84c) gehen auch aus den vorhergehenden Abschätzungen durch die Grenzübergänge  $\lambda_2 \uparrow 0$  bzw.  $\lambda_1 \downarrow 0$  hervor.

In Abb. 10 ist der Verlauf der Schranken (84) für  $|\varepsilon_e|$  im Falle des Rohverfahrens für einige Werte von  $\lambda_1$  und  $\lambda_2$  gezeichnet. Es wurden dabei die Annahmen  $\varepsilon = 0$ ,  $h = 1/10$  zugrunde gelegt und für die Komponenten  $q_v$  des Vektors  $Q$  in Tabelle 1 die Werte  $q_0 - q_1 = 1$  gewählt. Vergleichshalber sind noch die entsprechenden Schranken aus dem Satz II eingetragen.

### § 10. Asymptotischer Vergleich der drei Halbschrittverfahren

33. Es sollen für die Halbschrittverfahren in Tabelle 1 die Werte der hergeleiteten Fehlerschranken, die zum gleichen  $\varrho$  gehören, für  $\varrho \rightarrow \infty$  miteinander verglichen werden. Hierzu genügt offenbar ein Vergleich der entsprechenden Grössen  $t_{1,1}$  (für  $m = 1$ ) und  $t_{1,1}^2$  (für  $m = 2$ ), das heisst der Ausdrücke

$$t_{SV} = \left( \frac{h}{2} \lambda_1 + \sqrt{1 + \left( \frac{h}{2} \lambda_1 \right)^2} \right)^2, \quad (SV) \quad (85a)$$

$$t_{TV} = \frac{1 + \lambda_1 \operatorname{tg} \frac{h}{2}}{1 - \lambda_1 \operatorname{tg} \frac{h}{2}}, \quad (TV) \quad (85b)$$

$$t_{STV} = \left( \frac{\frac{1}{3} h \lambda_1 + \sqrt{1 + \left[ \left( \frac{1}{3} h \right)^2 - \left( \operatorname{tg} \frac{h/2}{3} \right)^2 \right] \lambda_1^2}}{1 - \lambda_1 \operatorname{tg} \frac{h/2}{3}} \right)^2. \quad (STV) \quad (85c)$$

Gemäss (61) schränken wir die Schrittlänge  $h$  durch

$$0 < \lambda_1 \operatorname{tg} \frac{h}{2} < 1 \quad (86)$$

ein.

34. **Satz III.** *Unter der Voraussetzung (86) gilt*

$$t_{SV} < t_{STV} < t_{TV}. \quad (87)$$

*Beweis.* a) Man vergrössert  $t_{STV}$ , wenn in (85c) die Ausdrücke  $h \lambda_1/3$  durch

$$\frac{2}{3} \lambda_1 \operatorname{tg} \frac{h}{2}$$

ersetzt werden. Mit

$$x = \frac{1}{3} \lambda_1 \operatorname{tg} \frac{h}{2}$$

ergibt sich die rechte Ungleichung in (87) aus der leicht zu verifizierenden Ungleichung

$$\left( \frac{2x + \sqrt{1+3x^2}}{1-x} \right)^2 < \frac{1+3x}{1-3x} \quad (x < 1).$$

b) Die Funktion

$$\varphi(t) = \frac{a + \sqrt{b-t^2}}{1-t} \quad (a > 0, b > 1)$$

ist wegen

$$(1-t)^2 \sqrt{b-t^2} \varphi'(t) = b-t + a \sqrt{b-t^2}$$

für  $0 < t < 1$  monoton wachsend. Daher verkleinert man  $t_{\text{STV}}$ , wenn man in (85c)  $\text{tg } h/2$  durch  $h/2$  ersetzt. Setzt man  $x = h \lambda_1/6$ , so ist die linke Ungleichung in (87) gesichert, sobald gezeigt ist, dass

$$3x + \sqrt{1+9x^2} < \frac{2x + \sqrt{1+3x^2}}{1-x} \quad (0 < x < 1) \quad (88)$$

gilt. Multiplikation beider Seiten von (88) mit  $1-x$  ergibt aber

$$3x^2 - (\sqrt{1+9x^2} - \sqrt{1+3x^2}) + x(\sqrt{1+9x^2} - 1) > 0,$$

und dies ist richtig, da

$$3x^2 > \sqrt{1+9x^2} - \sqrt{1+3x^2} = \frac{6x^2}{\sqrt{1+9x^2} + \sqrt{1+3x^2}}$$

gilt. Damit ist Satz III bewiesen.

### Literaturverzeichnis

- COLLATZ, L. (1951): *Numerische Behandlung von Differentialgleichungen* (Springer, Berlin; Reihe: Grundlehren der mathematischen Wissenschaften, Bd. 60).
- DOETSCH, G. (1921): *Über die graphische Integration von Differentialgleichungen erster Ordnung*. Z. angew. Math. Mech. **1**, 464–466.
- FRICKE, A. (1949): *Über die Fehlerabschätzung des Adamsschen Verfahrens zur Integration gewöhnlicher Differentialgleichungen 1. Ordnung*. Z. angew. Math. Mech. **29**, 165–178.
- GAUTSCHI, WALTER (1954): *Über eine Klasse von linearen Systemen mit konstanten Koeffizienten*. Wird erscheinen in den Comment. math. Helv.
- GRAMMEL, R. (1939): *Ein Gegenstück zum Meissnerschen Verfahren der graphischen Analysis*. Ing.-Arch. **10**, 395–411.
- KAMKE, E. (1944): *Differentialgleichungen, Lösungsmethoden und Lösungen*. Bd. 1, 3. Aufl. (Akademische Verlagsgesellschaft, Leipzig 1944).
- LUDWIG, R. (1951a): *Analytische Untersuchungen und konstruktive Erweiterungen zu den graphischen Integrationsverfahren von Meissner und Grammel*. Z. angew. Math. Mech. **31**, 120–130.
- (1951b): *Graphische Integrationsverfahren für die Meissnerschen Linienbilder bei Differentialgleichungen 1. Ordnung*. Z. angew. Math. Mech. **31**, 169–178.

- LUDWIG, R. (1951 c): *Graphische Integrationsverfahren für die Meissnerschen Linienbilder bei Differentialgleichungssystemen und Differentialgleichungen  $n$ -ter Ordnung*. Z. angew. Math. Mech. 31, 201–208.
- MEISSNER, E. (1931/32): *Graphische Analysis vermittelt des Linienbildes einer Funktion*. Schweiz. Bauztg. 98, 287–290, 333–335 (1931); 99, 27–30, 41–44, 67–69 (1932). (Auch als Sonderabdruck im Verlag der Schweiz. Bauztg. 1932 erschienen.)
- VON MISES, R. (1930): *Zur numerischen Integration von Differentialgleichungen*. Z. angew. Math. Mech. 10, 81–92.
- OSTROWSKI, A. (1939): *Sur la continuité relative des racines d'équations algébriques*. C. r. Acad. Sci. Paris 209, 777–779.
- (1937): *Über die Determinanten mit überwiegender Hauptdiagonale*. Comment. math. Helv. 10, 69–96.
- RICHTER, WILLY (1952): *Estimation de l'erreur commise dans la méthode de M. W. E. Milne pour l'intégration d'un système de  $n$  équations différentielles du premier ordre*. Bull. Soc. Neuchâteloise Sci. nat. 75, 6–43.
- SCHULZ, W. (1941): *Über das Meissnersche Integrationsverfahren für Differentialgleichungen erster Ordnung*. Deutsche Math. 6, 271–276.
- TOLLMEN, W. (1938): *Über die Fehlerabschätzung beim Adamsschen Verfahren zur Integration gewöhnlicher Differentialgleichungen*. Z. angew. Math. Mech. 18, 83–90.
- WEISSINGER, J. (1950): *Eine verschärfte Fehlerabschätzung zum Extrapolationsverfahren von Adams*. Z. angew. Math. Mech. 30, 356–363.

Manuskript eingegangen im Juli 1953.



# Über die zeichnerischen Ungenauigkeiten und die zweckmässige Bemessung der Schrittlänge beim graphischen Integrationsverfahren von Meissner-Ludwig<sup>1)</sup>

Mit 3 Abbildungen

Von Walter Gautschi, Basel

## § 1. Einleitung

1. Soll eine gewöhnliche Differentialgleichung beliebiger Ordnung  $n$ ,

$$y^{(n)} = f(x, y, \dots, y^{(n-1)}), \quad (1)$$

bei vorgegebenen Anfangswerten  $y^{(v)}(x_0) = y_0^{(v)}$  ( $v = 0, 1, \dots, n-1$ ) graphisch über ein gewisses  $x$ -Intervall  $J_x$  integriert werden, so eignet sich dazu in besonderem Masse ein Verfahren, das kürzlich von R. LUDWIG (1951)<sup>2)</sup> in Modifizierung des bekannten Meissnerschen Linienbildverfahrens [MEISSNER (1931/32)]<sup>3)</sup> entwickelt worden ist. Die einfachste Ausführung dieses Meissner-Ludwigschen Verfahrens, das sogenannte *Rohverfahren* (RV), ist ein Verfahren erster Ordnung, das heisst der Fehlerzuwachs bei Integration über einen Schritt der Länge  $h$  ist proportional  $h^2$ . Wesentlich bessere Resultate liefern das sogenannte *Sehnen-* (SV) und das *Tangentenverfahren* (TV), die beide von der Ordnung zwei sind, das heisst „Integrationsfehler“ proportional  $h^3$  haben. Durch eine geeignete Kombination des SV und des TV zum sogenannten *Sehnen-Tangentenverfahren* (STV) gelangt LUDWIG sogar zu einem Verfahren vierter Ordnung, zu einer Genauigkeit

<sup>1)</sup> Die vorliegende Arbeit entstand unter der Leitung von Herrn Prof. Dr. A. OSTROWSKI, der mit einigen wertvollen Anregungen die Niederschrift der Arbeit sehr gefördert hat.

<sup>2)</sup> Die Jahreszahlen in Klammern verweisen auf das Literaturverzeichnis am Schluss der Arbeit (S. 66).

<sup>3)</sup> Zum Verfahren von MEISSNER vgl. auch KAMKE (1944), S. 171 ff.

also, die bei graphischen Integrationsmethoden bisher kaum je erreicht worden ist.

2. Allen Verfahren ist gemeinsam, dass für jeden Integrationsschritt die Berechnung eines oder mehrerer Funktionswerte nötig ist. Es handelt sich also nicht um rein graphische, sondern um *graphisch-rechnerische* Näherungsverfahren. Bei der genaueren Analyse des Fehlers solcher Näherungsverfahren hat man prinzipiell zwei Arten von Fehler auseinanderzuhalten: *erstens* den sogenannten „*Verfahrensfehler*“, das heisst den Fehler, der das betreffende Näherungsverfahren aufweist, wenn sämtliche seine Ausführungsbestimmungen absolut genau durchgeführt werden; *zweitens* diejenigen Fehler, die auf *zeichnerische Ungenauigkeiten* und auf *Abrundungen* beim Ausführen der nötigen Rechnungen zurückgehen. Die Fehler der zweiten Art, die nicht für das Verfahren an sich, sondern nur für die Ausführungstechnik charakteristisch sind, wollen wir im folgenden kurz als „*technische Fehler*“ bezeichnen.

Beim Vergleich des Verfahrensfehlers mit den technischen Fehlern kann man sich auf einen einzigen Integrationsschritt beschränken, da sich beide Fehlerkomponenten in derselben Weise fortpflanzen [vgl. zum Beispiel (22) und die Bemerkungen in Nr. 14]. Wenn im folgenden kurz vom Verfahrensfehler und technischen Fehler einer Konstruktion die Rede ist, so ist darunter, wenn nicht ausdrücklich etwas anderes gesagt wird, der Zuwachs des entsprechenden Fehlers für einen *einzigsten Integrationsschritt* zu verstehen.

Ausserdem wird es im folgenden Zusammenhang nicht nötig sein, zwischen den Fehlern, mit denen die Lösungsfunktion und den Fehlern, mit denen ihre verschiedenen Ableitungen behaftet sind, zu unterscheiden.

3. Die nähere Untersuchung des *Verfahrensfehlers* zeigt [GAUTSCHI (1953)], dass für seine Grössenordnung die sogenannten *Quadraturfehler*  $R_v$  bestimmend sind [vgl. auch (22)]. Jedem der in Nr. 1 charakterisierten Integrationsverfahren ist ein System von  $n$  solchen Grössen  $R_v$  ( $v = 0, 1, \dots, n-1$ ) zugeordnet.

Ist  $y(x)$  die exakte, den Anfangsbedingungen genügende Lösungsfunktion der Differentialgleichung (1), so sei  $m_\mu$  eine obere Schranke der als stetig vorausgesetzten  $\mu$ -ten Ableitung von  $y(x)$  in  $J_x$ ,

$$|y^{(\mu)}(x)| \leq m_\mu \quad (x \in J_x; \mu = 0, 1, 2, \dots). \quad (2)$$

Dann lassen sich die Quadraturfehler  $R_v$  in den Fällen des RV, SV und des STV, die wir als Repräsentanten von Verfahren der

Ordnungen 1, 2 und 4 herausgreifen wollen, durch folgende Ausdrücke darstellen:

$$R_v^{(RV)} = \frac{h^2}{2} (m_v + 0,3 m_{v+1} + m_{v+2}), \quad (3,1)$$

$$R_v^{(SV)} = \frac{h^3}{12} (2 m_{v+1} + m_{v+3}), \quad (3,2)$$

$$R_v^{(STV)} = \frac{1}{90} \left( \frac{h}{2} \right)^5 (25 m_{v+1} + 10 m_{v+3} + m_{v+5})^2 \quad (3,3)$$

Es wird für unsere Zwecke genügen, als Masszahl für die Grössenordnung des Verfahrensfehlers dieser drei Verfahren die entsprechende Grösse  $R = \text{Max}_{0 < v < n-1} R_v$  zu verwenden (vgl. Nr. 15).

4. Man kann somit bei einem bestimmt vorgegebenen Differentialgleichungsproblem den Verfahrensfehler durch Wahl eines geeigneten Verfahrens und durch Verfügen über einen freien Konstruktionsparameter (die Schrittlänge  $h$ ) ohne weiteres beliebig verkleinern. Etwas Entsprechendes gilt für die *technischen Fehler* nur noch in sehr beschränktem Masse. Denn einmal sind die technischen Fehler für die verschiedenen Integrationsvarianten ungefähr gleich gross (vgl. § 3). Dann aber bestehen sie, abgesehen von den rechnerischen Ungenauigkeiten, zur Hauptsache in den unvermeidlichen *zeichnerischen Fehlern*, die wegen der endlichen Grösse der Strich- und Fleckenbreite (vgl. § 2) ein gewisses *Mindestmass nicht unterschreiten* können.

Zwar kann man versuchen, die technischen Fehler durch sorgfältiges und feines Zeichnen und vor allem durch die Wahl einer grossen Zeicheneinheit niedrig zu halten. Doch sind beiden Möglichkeiten aus praktischen Gründen gewisse Grenzen gesetzt; ausserdem nimmt die Zeichengenauigkeit nur *linear* mit der Erhöhung der Zeicheneinheit zu.

5. Unter diesen Umständen kommt bei graphischen Methoden der Bemessung der Schrittlänge  $h$ , oder, wie wir sagen wollen, der *Dimensionierung der Konstruktion*, eine besondere Bedeutung zu. Denn ist einmal die Zeicheneinheit fest gewählt<sup>3)</sup>, so hat man mit einer bestimmten Grössenordnung des technischen Fehlers zu rechnen. Die Konstruktion ist dann so zu dimensionieren, dass der Verfahrensfehler zum technischen Fehler in einem „vernünftigen Verhältnis“ steht.

<sup>1)</sup> Die Ausdrücke in (3) gelten unter der Bedingung  $0 < h \leq \pi/4$ , die in der Regel erfüllt ist. Etwas genauere Schranken, die an keine Bedingung über  $h$  gebunden sind, finden sich in meiner oben zitierten Arbeit: S. 415, 420, 422.

<sup>2)</sup> Im Falle des STV bezeichnet LUDWIG (1951)  $h/2$  als Schrittlänge.

<sup>3)</sup> Wir setzen als selbstverständlich voraus, dass die Zeicheneinheit so gross wie möglich gewählt wird.

An sich sind folgende drei Grenzfälle möglich:

A. der Verfahrensfehler ist wesentlich (das heisst etwa 100- bis 1000mal) kleiner als der technische Fehler;

B. der Verfahrensfehler übertrifft bei weitem den technischen Fehler;

C. der Verfahrensfehler ist ungefähr gleich dem technischen Fehler.

Wir sagen im Falle A, die Konstruktion sei *überdimensioniert* und im Falle B, sie sei *unterdimensioniert*.

6. Im Falle einer überdimensionierten Konstruktion sind zwar die Genauigkeitsverhältnisse günstig, da die Konstruktion praktisch nur technische Fehler aufweist. Indessen ist zu beachten, dass der Verfahrensfehler, und damit die Schrittlänge  $h$  vergrössert werden können, ohne dass sich der Gesamtfehler (pro Integrationsschritt) merklich ändert. Dies ist praktisch natürlich von besonderem Interesse, da die Vergrösserung der Schrittlänge in demselben Masse eine Verminderung der Schrittanzahl und somit auch eine Verminderung der Konstruktionsarbeit nach sich zieht. Ausserdem werden dadurch die technischen Fehler der *gesamten* Konstruktion entsprechend verkleinert.

Offenbar darf man den Schritt solange vergrössern, bis der Verfahrensfehler die Grössenordnung der technischen Fehler erreicht<sup>1)</sup>. Denn dadurch wird der Gesamtfehler (pro Integrationsschritt) höchstens ungefähr verdoppelt, die Anzahl der Schritte aber möglicherweise auf mehr als die Hälfte herabgesetzt.

Überdimensionierte Konstruktionen können sich leicht bei genaueren Integrationsverfahren, also beim SV und besonders beim STV einstellen.

Bei unterdimensionierten Konstruktionen, die also einen relativ grossen Verfahrensfehler aufweisen, sind die technischen Fehler derart unbedeutend, dass durch Verkleinerung der Schrittlänge die Genauigkeit der Konstruktion sicher gesteigert werden kann. Immerhin ist möglich, dass der Schritt dann praktisch zu klein wird; in diesem Fall ist es dann, wenn überhaupt Wert auf Genauigkeit gelegt wird, geboten, zu einem Verfahren mit kleinerem Verfahrensfehler überzugehen.

7. Beide Dimensionierungsmöglichkeiten A und B sind also, die eine aus praktischen Gründen, die andere aus Gründen der Genauigkeit, nicht anzustreben. Es bleibt daher nur die dritte Möglichkeit C, bei der tatsächlich weder die Vergrösserung noch

<sup>1)</sup> Die Schrittlänge darf allerdings nicht zu gross gewählt werden, weil im Falle des SV und STV die Konvergenz gewisser Iterationen gesichert werden muss (vgl. meine oben zitierte Arbeit, S. 419, 421).

die Verkleinerung der Schrittlänge eine wesentliche Genauigkeitssteigerung mehr ergibt.

Es ist daher zweckmässig, *die Konstruktion möglichst so zu dimensionieren, dass der Verfahrensfehler dem technischen Fehler ungefähr das Gleichgewicht hält.*

Durch diese Forderung wird auch bei graphischen Integrationsverfahren eine gewisse „*natürliche Schrittweite*“ definiert, wie sie COLLATZ (1942) im Falle numerischer Methoden eingeführt hat. Wählt man den Schritt kleiner als diese natürliche Schrittweite, so schafft man sich unnötig viel Konstruktions- und Rechenarbeit; wählt man ihn grösser als die natürliche Schrittweite, so holt man aus der betreffenden Konstruktion nicht das Maximum an Genauigkeit heraus.

8. Ziel der vorliegenden Arbeit wird sein, an Hand der Meissner-Ludwigschen Verfahren für die natürliche Schrittweite eine quantitative Bedingung anzugeben. Dies erfordert offenbar eine genauere Kenntnis der technischen Fehler, so dass es sich vor allem darum handeln wird, diese Fehler zunächst quantitativ zu analysieren.

Zum vornherein ist klar, dass eine vollständige Erfassung zeichnerischer Fehler wegen der Vielfalt der möglichen Ursachen nicht erreicht werden kann. Vielmehr wird man sich darauf beschränken müssen, das Ausmass dieser Fehler unter gewissen idealisierenden Annahmen abzuschätzen. Wir werden daher zunächst in § 2, im Anschluss an Gedankengänge von BÖHMER (1904), diese Annahmen treffen und die fundamentalen Forderungen, die an die Ausführung der Konstruktionen zu stellen sind, genau formulieren.

In § 3 werden wir dann in den Nrn. 12, 13 am Beispiel des Meissner-Ludwigschen Rohverfahrens die Konstruktion eingehend auf die technischen Fehler hin untersuchen. Es wird sich bei dieser Gelegenheit eine einfache Regel dafür ergeben, wie man die Fehlerschranken in meiner zitierten Arbeit für den Verfahrensfehler der gesamten Konstruktion (also mit Einschluss der Fortpflanzung) ergänzen kann, damit sie auch mit Einbezug der technischen Fehler gültig bleiben. Entsprechende Regeln lassen sich auch für die übrigen Integrationsverfahren herleiten und werden in der Nr. 14 zusammengestellt.

In der Nr. 15 werden dann die technischen Fehler dem Verfahrensfehler gegenübergestellt. Es werden sich durch die Formeln (18) für jede einzelne Integrationsvariante und für jedes beliebige Differentialgleichungsproblem (1) gewisse Masszahlen  $u$  einführen lassen, mit deren Hilfe man feststellen kann, welchen

Anteil am Gesamtfehler die technischen Fehler in dem betreffenden Falle ungefähr haben. Die Bedingung für zweckmässige Dimensionierung lautet dann, dass  $u$  ungefähr gleich 1 sein soll (Dimensionierungsprinzip in Nr. 15). Da  $u$  von der Schrittlänge  $h$  abhängt, kann dieser Bedingung die natürliche Schrittweite entnommen werden.

In § 4 wird am Beispiel (20) einer Differentialgleichung zweiter Ordnung zunächst eine vollständige Fehlerabschätzung durchgeführt und dann insbesondere der Einfluss der technischen Fehler verfolgt. Es wird sich bei diesem Beispiel als zweckmässig erweisen, im Falle des STV die Schrittlänge ganz erheblich zu vergrössern.

## § 2. Bezeichnungen und Voraussetzungen

9. Wir werden in diesem Zusammenhang das Abbild einer idealen Geraden nach BÖHMER (1904) als *Strich* bezeichnen und das Abbild eines idealen Punktes als *Fleck*. Wir nehmen an, dass jeder Strich so beschaffen ist, dass genau ein Parallelstreifen von minimaler Breite  $b$  existiert, der den Strich vollständig im Innern enthält. Es gelte stets  $b \leq 2\omega_1$ , wo  $\omega_1$  eine im folgenden feste Konstante bedeutet, die die Feinheit der Zeichnung messen soll<sup>1)</sup>. Als *Mittellinie eines Striches* bezeichnen wir die Mittellinie des zugehörigen Minimalstreifens. Analog setzen wir voraus, dass zu jedem Fleck eindeutig eine Kreisscheibe von minimalem Radius  $r$  gehört, die den Fleck vollständig überdeckt und dass stets  $r \leq \omega_2$  ist, wo über  $\omega_2$  das über  $\omega_1$  Gesagte gilt (vgl. Fussnote <sup>1)</sup>). Den *Mittelpunkt des Flecks* definieren wir als Zentrum der entsprechenden Kreisscheibe.

Den Grössen  $2\omega_1$  und  $2\omega_2$  sind durch die Bedingung, dass Strich und Fleck sichtbar sein sollen, gewisse untere Schranken gesetzt. Für das Zeichnen mit Bleistift oder Reissfeder hat BÖHMER (1904) die Grössen

$$2\omega_1 = 0,08 \text{ mm}, \quad 2\omega_2 = 0,10 \text{ mm} \quad (4)$$

gefunden. Mit diesen Werten wird nicht nur der Forderung der Sichtbarkeit genügt, sondern auch ein längeres Konstruieren ohne Ermüdung der Augen gewährleistet.

Einen Strich heissen wir mit einem Fleck in *inzidenter Lage*, wenn die Mittellinie des Striches vom Mittelpunkt des Flecks einen Abstand  $\leq \omega_0$ ,  $\omega_0 = \omega_1 + \omega_2$  hat (vgl. Abb. 1).

<sup>1)</sup> Die Grösse der Konstante  $\omega_1$  hängt, abgesehen von der Erfahrung und Geschicklichkeit des Zeichners, in erster Linie von der Güte der Zeicheninstrumente und von der Körnigkeit des Zeichenpapiers ab.

10. Zur praktischen Durchführung des Meissner-Ludwigschen Verfahrens sind folgende Elementaroperationen wiederholt anzuwenden:

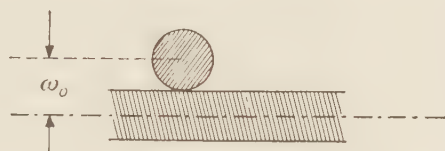


Abb. 1. Inzidente Lage von Fleck und Strich.

- I. Markieren eines Punktes auf einer Geraden.
- II. Ziehen einer Geraden durch einen Punkt.
- III. Abtragen einer Zahl als Strecke auf einer Geraden.
- IV. Messen des Abstandes zweier Punkte.

Wir stellen an die Ausführung dieser Operationen die folgenden *Minimalforderungen*. Es soll im Resultat beziehungsweise

- I<sup>0</sup>. der Markierungsfleck mit dem vorgegebenen Strich inzident liegen;
- II<sup>0</sup>. der Strich mit dem vorgegebenen Fleck inzident liegen;
- III<sup>0</sup>. die Distanz vom Mittelpunkt des einen „Endflecks“ der Strecke zum Mittelpunkt des anderen sich von der vorgegebenen Zahl höchstens um eine Konstante  $\eta$  unterscheiden;
- IV<sup>0</sup>. das Messergebnis und der Abstand der Zentren der beiden vorgegebenen Flecken sich höchstens um  $\eta$  unterscheiden.

Die Genauigkeit der rein *geometrisch-konstruktiven* Operationen I, II wird somit durch die Konstante  $\omega_0$  gekennzeichnet, die Genauigkeit der *metrischen* Operationen III, IV durch die weitere Konstante  $\eta$ . Im Prinzip können zwar die metrischen Operationen in geometrisch-konstruktive zerlegt werden, zum Beispiel das Abtragen einer Zahl durch Markieren der beiden Spitzen eines Stechzirkels auf den entsprechenden Teilungsstrichen des Maßstabes. Dies würde für  $\eta$  den Wert  $2\omega_0$  nahelegen. Indessen bringt der Gebrauch des Maßstabes weitere Ungenauigkeiten mit sich, so zum Beispiel die Ungenauigkeit der Skala, und vor allem solche, die durch das Schätzen zwischen zwei Skalenteilen entstehen. Wir haben daher für die metrischen Konstruktionen eine unabhängige Konstante  $\eta$  eingeführt, die praktisch etwas grösser als  $2\omega_0$  zu wählen ist.

Soll insbesondere der Schnittpunkt zweier Geraden markiert werden, so verlangen wir gemäss  $I^0$ , dass der Markierungsfleck mit *beiden* Strichen inzident liegt<sup>1)</sup>. Dual hierzu soll der Verbindungsstrich zweier Flecken mit beiden Flecken inzident liegen. Ist nur die Richtung der Geraden durch einen Punkt vorgeschrieben, so setzen wir voraus, dass die Mittellinie des entsprechenden Striches genau in vorgeschriebener Richtung verläuft und der Strich selbst mit dem Fleck inzident liegt. Diese Festsetzung wird durch den Umstand nahegelegt, dass man eine gegebene Richtung praktisch sehr genau reproduzieren kann.

Die Inzidenzforderungen  $I^0$ ,  $II^0$  werden bei gleicher Feinheit der Zeichnung von verschiedenen Zeichnern, entsprechend ihrer Übung und Sorgfalt, in verschieden hohem Masse erfüllt. Dem kann dadurch Rechnung getragen werden, dass man bei der Definition der Inzidenz die Konstante  $\omega_0$  durch eine *individuelle* Konstante  $\omega < \omega_0$  ersetzt und  $\eta$  entsprechend verkleinert.

### § 3. Diskussion der technischen Fehler

11. Wir führen der Einfachheit halber die Untersuchung zuerst explizit am Beispiel des RV für  $n = 1$  durch. Die Übertragung auf  $n > 1$  sowie auf das SV und STV ist derart naheliegend, dass wir in diesen Fällen uns auf die Angabe der entsprechenden Resultate beschränken können.

Wir schicken zunächst eine kurze Beschreibung des Meissner-Ludwigschen RV für  $n = 1$  voraus.

Es sei die Schrittlänge  $h > 0$  festgelegt, also die  $x$ -Werte

$$x_\varrho = x_0 + \varrho h \quad (\varrho = 1, 2, \dots), \quad (5)$$

in denen Näherungen  $y_\varrho, y'_\varrho$  der Funktionswerte  $y(x_\varrho)$  bzw.  $y'(x_\varrho)$  gefunden werden sollen.  $y(x)$  ist die exakte Lösung der Differentialgleichung (1) für  $n = 1$  mit  $y(x_0) = y_0$ .

Für das Verfahren von MEISSNER ist die Verwendung von *Polarkoordinaten* charakteristisch. Sei  $O$  der Pol und der von  $O$  ausgehende Halbstrahl  $s_0$  unsere Bezugsachse. Wir fassen  $h$  als Bogenmass eines Winkels auf und bezeichnen die Halbstrahlen, die von  $O$  ausgehen und mit  $s_0$  den positiven Winkel  $\varrho h$  bilden, mit  $s_\varrho$ . Die Strahlen  $s'_\varrho$  ( $\varrho = 0, 1, 2, \dots$ ) mögen aus  $s_\varrho$  durch Drehung im positiven Sinn um den Winkel  $\pi/2$  entstehen.  $s_\varrho, s'_\varrho$  werden als Träger der Näherungspunkte  $M_\varrho$  bzw.  $M'_\varrho$  dienen, derart, dass  $OM_\varrho = y_\varrho, OM'_\varrho = y'_\varrho$  ist. (Positive Werte von  $y_\varrho$  sind dabei

<sup>1)</sup> An sich könnte man auch fordern, dass der Mittelpunkt des Markierungsflecks im gemeinsamen Parallelogramm der beiden Striche liegt.

in Richtung von  $s_0$ , negative Werte auf der rückwärtigen Verlängerung von  $s_0$  gelegen, und dasselbe gilt für  $y'_0, s'_0$ .)

Zunächst findet man  $M_0$ , indem man  $y_0$  unter Beachtung des Vorzeichens von  $O$  aus auf  $s_0$  abträgt, und  $M'_0$ , indem man  $y'_0 - f(x_0, y_0)$  auf  $s'_0$  abträgt. Ergänzt man  $M_0 O M'_0$  zum Rechteck  $M_0 O M'_0 \mathfrak{M}_0$ , so definieren wir den ersten Näherungspunkt  $M_1$  als Fusspunkt des Lotes von  $\mathfrak{M}_0$  auf  $s_1$  (vgl. Abb. 2). Misst man  $OM_1 = y_1$  aus und berechnet man  $y'_1$  durch  $y'_1 = f(x_1, y_1)$ , so kann weiter  $M'_1$  auf  $s'_1$  aufgetragen werden und vom Eckpunkt  $\mathfrak{M}_1$  des Rechtecks  $M_1 O M'_1 \mathfrak{M}_1$  das Lot auf  $s_2$  gefällt werden. Dies liefert  $OM_2 = y_2$  usw.

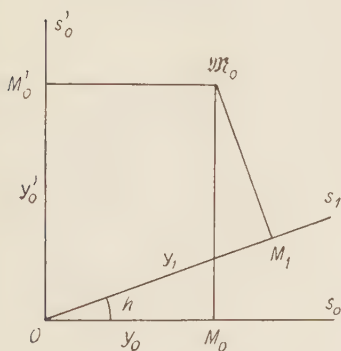


Abb. 2.

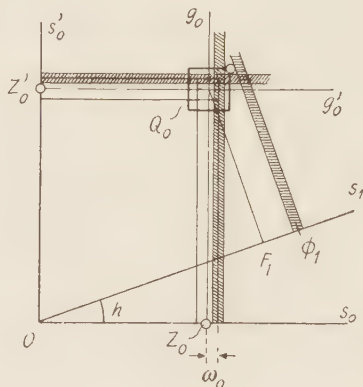


Abb. 3.

Abb. 2. Das Meissner-Ludwigsche Rohverfahren.

Abb. 3. Das Meissner-Ludwigsche Rohverfahren bei begrenzter Zeichengenauigkeit.

Die Folge der Näherungswerte  $y_0, y'_0$ , die man so erhält, ist, absolut genaues Zeichnen und Rechnen vorausgesetzt, durch die Gleichungen

$$y_{e+1} = y_e \cosh h + y'_e \sinh h, \quad y'_e = f(x_e, y_e) = f_e \quad (e = 0, 1, 2, \dots) \quad (6)$$

gegeben. Ändert man  $y_0$  und  $y'_0$  beziehungsweise um Größen  $\Delta y_0 = a, \Delta y'_0 = a'$ , so ändert sich  $y_1$  um

$$\Delta y_1 = a \cosh h + a' \sinh h. \quad (7)$$

12. Wir untersuchen nun, inwiefern (6) abzuändern ist, wenn nicht mehr absolute Genauigkeit der Zeichnung gefordert wird, hingegen angenommen wird, dass die Voraussetzungen des § 2

erfüllt sind und ausserdem ein genaues Polarkoordinatennetz<sup>1)</sup> vorliegt. Wir zerlegen zu diesem Zweck die Konstruktion in ihre einzelnen Schritte (vgl. Abb. 3).

a) Es werden die Punkte  $M_0, M'_0$  auf den Strahlen  $s_0, s'_0$  abgetragen. Die Mittelpunkte  $Z_0, Z'_0$  der zugehörigen Flecken haben nach III<sup>0</sup> von den exakten Punkten  $M_0, M'_0$  je einen Abstand  $\leq \eta$ . (Wir können dabei annehmen, dass  $Z_0, Z'_0$  genau auf den Strahlen  $s_0$  bzw.  $s'_0$  liegen, da eine Verrückung senkrecht zu diesen Strahlen für das folgende belanglos ist.)

Es bezeichne  $g_0^{(v)}$  ( $v = 0, 1$ ) die Geraden durch  $Z_0^{(v)}$  senkrecht zu  $s_0^{(v)}$ , und  $F_1$  den Fusspunkt des Lotes vom Schnittpunkt dieser Geraden auf  $s_1$ . Ist  $M_1$  der Näherungspunkt gemäss der Abb. 2, so gilt wegen (7) für die Distanz von  $F_1$  nach  $M_1$ , wenn  $\cosh h + \sinh h$  durch 1 angenähert wird,

$$|F_1 M_1| \leq \eta. \quad (8)$$

b) Durch die Flecken in  $Z_0, Z'_0$  sind Striche zu ziehen parallel zu den Geraden  $g_0$  bzw.  $g'_0$ . Nach II<sup>0</sup> liegen die zugehörigen Mittellinien sicher in den Streifen  $S_0, S'_0$  der Breite  $2\omega_0$  längs den Geraden  $g_0, g'_0$ . Es bezeichne  $D_0$  den Durchschnitt  $D_0 = S_0 S'_0$ .

c) Der Schnittpunkt der eben gezogenen Geraden wird durch einen Fleck markiert. Der Mittelpunkt dieses Flecks ist wegen I<sup>0</sup> sicher im Quadrat  $Q_0$  enthalten, das aus dem Quadrat  $D_0$  durch Ränderung um den Betrag  $\omega_0$  entsteht.

d) Durch den Markierungspunkt ist die Projizierende  $p_1$  auf  $s_1$  zu ziehen. Die Mittellinie dieser Projizierenden steht nach II<sup>0</sup> vom Quadrat  $Q_0$  in Richtung  $s_1$  höchstens um den Betrag  $\omega_0$  ab. Der Fusspunkt  $\Phi_1$  dieser Mittellinie auf  $s_1$  hat daher vom Punkt  $F_1$  höchstens eine Entfernung  $\omega_0 + 2\omega_0(\cosh h + \sinh h)$  oder, wenn wieder  $h$  gegen 1 vernachlässigt wird, eine Entfernung

$$|F_1 \Phi_1| \leq 3\omega_0. \quad (9)$$

e) Der Näherungswert  $y_1$  entsteht durch Ausmessen der Entfernung von  $O$  nach  $\Phi_1$ , was nach IV<sup>0</sup> mit einem weiteren Fehlerzuwachs  $\leq \eta$  verbunden ist. Aus (8) und (9) folgt  $|\Phi_1 M_1| \leq 3\omega_0 + \eta$  und somit für die Näherung  $y_1$

$$y_1 = y_0 \cosh h + y'_0 \sinh h + \langle 3\omega_0 + 2\eta \rangle, \quad (10)$$

<sup>1)</sup> Die Ungenauigkeiten des Polarkoordinatensystems, insbesondere die Winkelfehler der Strahlen  $s_\alpha$ , liefern natürlich auch einen Beitrag zum Gesamtfehler der Konstruktion. Sie können sogar leicht die Grösse der technischen Fehler erreichen, wenn die Strahlen  $s_\alpha$  nicht mit der nötigen Sorgfalt gezeichnet werden. Bei vorgedrucktem Polarkoordinatenpapier wäre es wünschenswert, wenn der Hersteller über die Genauigkeit der Winkel bestimmte Angaben machen könnte.

wenn allgemein für ein  $a > 0$  unter  $\langle a \rangle$  eine Grösse verstanden wird, die absolut höchstens gleich  $a$  ist. In (10) sind die Konstanten  $\omega_0$  und  $\eta$  in der Zeicheneinheit auszudrücken, die der Konstruktion zugrunde liegt. Beträgt diese Einheit  $e$  mm, so ist also zum Beispiel für  $\omega_0$  etwa nach (4) zu setzen:  $\omega_0 = 0,09/e$ .

f)  $y'_1$  wird gemäss (6) mit Hilfe des Wertes (10) berechnet. Unter der Annahme, dass die Rechnungen bis auf  $d\%$  genau ausgeführt werden, gilt demnach

$$y'_1 = (1 + \langle \delta \rangle) f(x_1, y_1), \quad \delta = \frac{d}{100}.$$

Benutzt man einen Rechenschieber von üblicher Länge, so kann etwa  $\delta = 3 \cdot 10^{-3}$  gesetzt werden.

Für  $\varrho \geq 1$  sind die obigen Überlegungen nur insofern zu ändern, als der Fehleranteil  $\omega_0$ , der beim Zeichnen der Geraden  $g_0$  zu veranschlagen war, für die Gerade  $g_\varrho$  in Wegfall kommt, da  $g_\varrho$  mit der bereits gezeichnet vorliegenden Projizierenden  $p_\varrho$  identisch ist. Aus Gründen der Einfachheit machen wir jedoch von dieser sich bietenden Verschärfungsmöglichkeit keinen Gebrauch.

Wir schreiben daher für allgemeines  $\varrho$ , indem wir noch  $\omega_0$  durch die individuelle Konstante  $\omega$  ersetzen, an Stelle von (6)

$$y_{\varrho+1} = y_\varrho \cosh + y'_\varrho \sinh + \langle 3\omega + 2\eta \rangle (\varrho = 0, 1, 2, \dots),$$

$$y'_\varrho = (1 + \langle \delta \rangle) f_\varrho, \quad f_\varrho = f(x_\varrho, y_\varrho). \quad (11)$$

Die Werte von  $y_\varrho$ ,  $y'_\varrho$  sind wohlbestimmte Zahlen, nämlich die Ergebnisse der jeweiligen Messungen und Rechnungen, und werden während der Konstruktion laufend tabelliert.

Das Zusatzglied in (11) ist so bemessen, dass auch der ungünstigste, in Abb. 3 dargestellte Fall berücksichtigt wird, in dem sämtliche konstruktiven Ungenauigkeiten sich in derselben Richtung auswirken. Praktisch tritt dieser Fall kaum je auf, umso mehr, als es der Zeichner in der Hand hat, eine allfällig begangene Ungenauigkeit in der Folge durch entsprechende Korrekturen auszugleichen. Der im folgenden geschätzte Zeichenfehler wird daher den tatsächlich begangenen um ein Vielfaches übertreffen können.

13. Gemäss der Bedeutung des Quadraturfehlers  $R_0^{(RV)}$  gilt für die exakte Lösungsfunktion  $y(x)$  die Relation

$$y(x_{\varrho+1}) = y(x_\varrho) \cosh + y'(x_\varrho) \sinh + \langle R_0^{(RV)} \rangle. \quad (12)$$

Bezeichnet man mit  $\varepsilon_\varrho$  den Fehler  $\varepsilon_\varrho = y_\varrho - y(x_\varrho)$ , so erhält man durch Subtraktion der Gleichung (12) von (11) für  $\varepsilon_\varrho$  die Abschätzung

$$|\varepsilon_{\varrho+1}| \leq |\varepsilon_\varrho| + h |(1 + \langle \delta \rangle) f_\varrho - f[x_\varrho, y(x_\varrho)]| + R_0^{(RV)} + 3\omega + 2\eta. \quad (13)$$

In einem konvexen Bereich  $B$  der  $(x, y)$ -Ebene, der alle Punkte  $(x_\varrho, y_\varrho)$ ,  $[x_\varrho, y(x_\varrho)]$  enthält, möge  $\partial f / \partial y$  stetig sein. Dann genügt  $f$  in  $B$  einer Lipschitz-Bedingung, woraus insbesondere

$$|f(x_\varrho, y_\varrho) - f[x_\varrho, y(x_\varrho)]| \leq k |y_\varrho - y(x_\varrho)| = k |\varepsilon_\varrho|$$

folgt, mit der Lipschitz-Konstanten  $k$ . Sei ferner

$$|f_\varrho| \leq \mu \quad (\varrho = 0, 1, 2, \dots);$$

dann ergibt (13) schliesslich die rekursive Abschätzung

$$|\varepsilon_{\varrho+1}| \leq (1 + h k) |\varepsilon_\varrho| + R_0^{(RV)} + 3\omega + 2\eta + \delta \mu h. \quad (14)$$

14. Offenbar erhält man aus (14) die entsprechende Abschätzung für den reinen Verfahrensfehler, indem man  $\omega = \eta = \delta = 0$  setzt. Jede independente Abschätzung des Verfahrensfehlers, die man in diesem Falle aus (14) herleitet, kann daher sofort auf den Fall endlicher technischer Fehler umgeschrieben werden, indem man in den Schranken für  $\varepsilon_\varrho$  überall

$$R_0^{(RV)} \quad \text{durch} \quad R_0^{(RV)} + 3\omega + 2\eta + \delta \mu h$$

ersetzt.

Ein ganz entsprechender Sachverhalt besteht auch für die übrigen Integrationsvarianten und für  $n > 1$ . Für  $\mu$  ist dann das Maximum

$$\mu = \text{Max}_\varrho |f_\varrho|, \quad f_\varrho = f(x_\varrho, y_\varrho, \dots, y_\varrho^{(n-1)}) \quad (15)$$

zu nehmen, wo  $y_\varrho, \dots, y_\varrho^{(n-1)}$  die Näherungswerte für  $y(x)$  und seine Ableitungen an den Stellen (5) bedeuten.

Das allgemeine Resultat lässt sich dahin zusammenfassen, dass (unter den Annahmen des § 2) sämtliche in meiner zitierten Arbeit hergeleiteten Fehlerschranken auch mit Einbezug der technischen Fehler gültig bleiben, wenn in den Schranken die Quadraturfehler  $R_\nu$  wie folgt ersetzt werden:

$$R_\nu \rightarrow R_\nu + Z \quad (\nu = 0, 1, \dots, n-2), \quad R_{n-1} \rightarrow R_{n-1} + Z + \delta \mu h, \quad (16)$$

wo  $Z$  durch

$$Z = \begin{cases} 3 \omega + 2 \eta & (\text{RV}) \\ 5 \omega + 2 \eta & (\text{SV}) \\ 4 \omega + 2 \eta & (\text{STV}) \end{cases} \quad (17)$$

und  $\mu$  durch (15) definiert sind.

15. Wie schon in der Einleitung erwähnt wurde, sind die Verfahrensfehler (pro Integrationsschritt) von der Grössenordnung der Quadraturfehler  $R_v$ . Da sich die technischen Fehler in einer Vergrösserung dieser Quadraturfehler gemäss (16) und (17) auswirken, können nun die in Nr. 5 der Einleitung beschriebenen Dimensionierungsverhältnisse mit Hilfe der Grössen

$$R = \text{Max}_{0 \leq v < n} R_v \quad \text{und} \quad T = Z + \delta \mu h$$

charakterisiert werden. Offenbar ist die Konstruktion dann überdimensioniert, wenn  $T$  wesentlich grösser als  $R$  ist, und unterdimensioniert, wenn  $T$  wesentlich kleiner ist als  $R$ . Zur Beurteilung der Dimensionierung einer Konstruktion ist es daher zweckmässig, die Verhältnissgrösse

$$u = \frac{Z + \delta \mu h}{\text{Max}_v R_v},$$

die wir als *Dimensionierungszahl* bezeichnen wollen, einzuführen.

Es muss beachtet werden, dass diese Masszahl davon abhängt, wie gut die Konstanten  $m_\mu$ , die in  $R_v$  stecken, abgeschätzt worden sind. Sie entspricht nur dann dem wirklichen Verhältnis der technischen Fehler zum Verfahrensfehler, wenn die *geschätzten* Grössen  $Z + \delta \mu h$  und  $R_v$  die *tatsächlich begangenen* technischen Fehler bzw. Quadraturfehler prozentual ungefähr um denselben Betrag übertreffen<sup>1)</sup>.

Für die Dimensionierungszahlen der in § 1 betrachteten Integrationsvarianten ergeben sich nach (3) und (17) die folgenden Ausdrücke:

$$u_{\text{RV}} = \frac{3 \omega + 2 \eta + \delta \mu h}{h^2/2 \text{Max}_{0 \leq v < n} (m_v + 0,3 m_{v+1} + m_{v+2})}, \quad (18,1)$$

$$u_{\text{SV}} = \frac{5 \omega + 2 \eta + \delta \mu h}{h^3/12 \text{Max}_{0 \leq v < n} (2 m_{v+1} + m_{v+3})}, \quad (18,2)$$

$$u_{\text{STV}} = \frac{4 \omega + 2 \eta + \delta \mu h}{1/90 (h/2)^5 \text{Max}_{0 \leq v < n} (25 m_{v+1} + 10 m_{v+3} + m_{v+5})}. \quad (18,3)$$

<sup>1)</sup> Die geschätzten technischen Fehler werden in der Regel ungefähr das Zehnfache der wirklich begangenen technischen Fehler betragen. Demgegenüber lassen sich die Quadraturfehler, falls nicht zu hohe Ableitungen von  $y$  auftreten, möglicherweise etwas genauer abschätzen. Sobald allerdings höhere Ableitungen auftreten (wie zum Beispiel beim STV), so ist auch dann die Abschätzung meist nur bis auf einen Faktor 10 möglich.

Wir formulieren nunmehr die in Nr. 7 der Einleitung als zweckmässig erkannte Dimensionierungsforderung quantitativ in dem folgenden

**Dimensionierungsprinzip.** *Jede der Meissner-Ludwigschen Konstruktionen (RV, SV, STV) ist möglichst so anzulegen, dass für die entsprechende Dimensionierungszahl  $u$  in (18) ungefähr*

$$u = 1 \quad (19)$$

*gilt.*

Zur Berechnung der Dimensionierungszahlen in (18) sind in erster Linie die Schranken  $m_\mu$  aus (2) für die Lösungsfunktion und einige ihrer Ableitungen zu bestimmen. Dies ist in der Regel der schwierigste Teil der Rechnung. Man wird versuchen müssen, will man keine Daten benutzen, die vom Ablauf der Konstruktion abhängen, durch mehrmaliges Differenzieren der gegebenen Differentialgleichung die Ableitungen abzuschätzen<sup>1)</sup>. Der Umstand, dass die Schranken nicht besonders genau zu sein brauchen, kann diese Abschätzung gelegentlich erleichtern.

Die Parameter  $\omega$ ,  $\eta$  ergeben sich gemäss § 2, sobald die Zeicheneinheit festgelegt ist, während für  $\delta$ , falls mit dem Rechenschieber gerechnet wird,  $\delta = 3 \cdot 10^{-3}$  zu setzen ist. Für  $\mu$  kann natürlich an Stelle von (15) eine Schranke der Funktion  $f(x, y, \dots, y^{(n-1)})$  in einem „hinreichend grossen“ Bereich des  $(x, y, \dots, y^{(n-1)})$ -Raumes genommen werden.

#### § 4. Beispiel

16. Wir wählen als Beispiel die [bereits von GRAMMEL (1939) graphisch behandelte] Differentialgleichung zweiter Ordnung für die Hängekurve eines Kabels,

$$2 y'' = \sqrt{1 + y'^2} + 1, \quad (20)$$

mit den Anfangsbedingungen

$$y(0) = 1, \quad y'(0) = 0. \quad (20^0)$$

Gesucht sei der Funktionswert  $y(\pi/2)$ .

Mit den Konstruktionsdaten  $e = 50 \text{ mm}^2$ ,  $h = \pi/18 = 0,1745 \dots$  lieferte das Meissner-Ludwigsche Verfahren folgende Resultate:

$$\begin{aligned} y_g &= 1,945; & y'_g &= 1,69; & y''_g &= f_g = 1,48 & (\text{RV}) \\ y_g &= 2,37; & y'_g &= 1,865; & y''_g &= f_g = 1,557 & (\text{SV}). \end{aligned} \quad (21)$$

<sup>1)</sup> Will man hinterher, nach Ausführung der Konstruktion, die Dimensionierung beurteilen, oder ist orientierungshalber bereits das RV angewandt worden, so kann auch das Differenzenschema der gefundenen  $f_v$ -Werte verwendet werden. Man wird dann an Stelle der höheren Ableitungen von  $f$  die höheren Differenzen abschätzen.

<sup>2)</sup> Bei der Wahl der Einheit  $e$  musste auf das relativ kleine Format der vorgedruckten Koordinatenblätter Rücksicht genommen werden.

Bei der gewählten Dimensionierung der Konstruktion kam das STV nicht in Betracht, da die Werte des SV und des TV innerhalb der Zeichengenauigkeit nicht zu unterscheiden waren. Alle Rechnungen wurden mit dem Rechenschieber ausgeführt.

Die exakte Lösung von (20), (20<sup>0</sup>) lautet in Parameterdarstellung<sup>1)</sup>

$$x = 2 \lg \frac{1+t}{1-t} - 2t, \quad (|t| < 1).$$

$$y = 2 \lg (1 - t^2) + \frac{1+3t^2}{1-t^2}$$

Insbesondere ist  $y(\pi/2) = 2,3541$ ,  $y'(\pi/2) = 1,8722$ .

17. Gemäss (1) ist für  $f$  zu setzen

$$f(x, y, y') = \frac{1}{2} (1 + \sqrt{1 + y'^2}).$$

Für die Fälle, in denen  $f$ , wie hier, von  $y$  unabhängig ist, sind in meiner zitierten Arbeit für den Verfahrensfehler besondere Schranken hergeleitet worden, die zeigen, dass in diesen Fällen die Fortpflanzung des Fehlers günstiger verläuft, als wenn noch  $y$  in  $f$  auftreten würde. Bezeichnet  $k_1$  das Maximum

$$k_1 = \max_B \left| \frac{\partial f}{\partial y'} \right|,$$

erstreckt über einen hinreichend grossen Bereich  $B$  des  $(x, y, y')$ -Raumes, so gelten für die Verfahrensfehler  $\varepsilon_e = y_e - y(x_e)$  und  $\varepsilon'_e = y'_e - y'(x_e)$

( $x_e = q \pi/18$ ) die Abschätzungen<sup>2)</sup>

$$\begin{aligned} |\varepsilon_e| &\leq \frac{R_1}{h k_1^2} (t^e - 1) + \frac{k_1 R_0 - R_1}{k_1} q + \frac{\varepsilon}{k_1} (t^e - 1) + \varepsilon, \\ |\varepsilon'_e| &\leq \frac{R_1}{h k_1} (t^e - 1) + \varepsilon t^e. \end{aligned} \quad (22)$$

Hier ist  $\varepsilon$  eine Schranke für eventuelle durch Abrundungen entstandene Fehler in den Anfangswerten  $y_0, y'_0$ . In unserem Beispiel kann, da keine solche Abrundungen vorzunehmen sind,  $\varepsilon = 0$  gesetzt werden.  $R_0, R_1$  sind die Quadraturfehler (3), während  $t$  für die verschiedenen Integrationsvarianten die folgende Bedeutung hat:

$$t = \begin{cases} 1 + h k_1 & \text{(RV)} \\ \frac{1 + k_1 \operatorname{tg} h/2}{1 - k_1 \operatorname{tg} h/2} & \text{(SV)} \\ \left( \frac{h/3 k_1 + \sqrt{1 + (h^2 - \operatorname{tg}^2 h/2) k_1^2/9}}{1 - (k_1/3) \operatorname{tg} h/2} \right)^2 & \text{(STV).} \end{cases}$$

In unserem Beispiel ist, wenn für  $B$  der Bereich  $0 \leq x \leq \pi/2$ ,  $-\infty \leq y, y' \leq \infty$  genommen wird,

$$k_1 = \frac{1}{2} \sup_B \left| \frac{y'}{\sqrt{1 + y'^2}} \right| = \frac{1}{2}.$$

<sup>1)</sup> Vgl. KAMKE (1944), S. 556, wo der Parameter  $u$  verwendet wird, der mit unserem Parameter  $t$  vermöge  $t = \operatorname{tg} u/2$  zusammenhängt.

<sup>2)</sup> GAUTSCHI (1953), S. 432, Formel (84b).

Die Werte von  $t$  errechnen sich daraus beziehungsweise zu 1,088; 1,092; 1,091.

18. Zur Bestimmung der Quadraturfehler  $R_\nu$  nach (3) sind zunächst die Schranken  $m_\mu$  aus (2) zu bestimmen. Diese lassen sich durch fortgesetzte Differentiation von (20) nach  $x$  durch  $m_1$  ausdrücken. Die Rechnung ergibt

$$m_2 = \frac{1}{2} (1 + \sqrt{1 + m_1^2}), \quad m_3 = \frac{1}{2} m_2, \quad m_4 = \frac{1}{2} (1 + m_3), \quad m_5 = \frac{1}{4} (1 + 2 m_4),$$

$$m_6 = \frac{1}{2} (1 + m_5). \quad (23)$$

Um noch  $m_1$  und  $m_0$  zu bestimmen, benutzen wir die Darstellung

$$2 y'(x) = x + \int_0^x \sqrt{1 + y'^2(\xi)} \, d\xi, \quad (24)$$

aus der hervorgeht, dass  $y'(x)$  mit  $x$  monoton wächst. Setzt man  $m = y'(\pi/2)$ , so entnimmt man (24) die Abschätzung

$$2 m < \frac{\pi}{2} + \frac{\pi}{2} \sqrt{1 + m^2},$$

woraus

$$m < \frac{8\pi}{16 - \pi^2} = 4,10$$

folgt. Wir setzen daher  $m_1 = 4,10$  und für  $m_0$  den Wert

$$m_0 = 1 + (\pi/2) m_1 = 7,45^1).$$

Damit ergeben sich aus (23) für die Konstanten  $m_\mu$  die Zahlenwerte

$$m_0 = 7,45, \quad m_1 = 4,10, \quad m_2 = 2,61, \quad m_3 = 1,31,$$

$$m_4 = 1,16, \quad m_5 = 0,83, \quad m_6 = 0,915. \quad (25)$$

Setzt man diese Werte in (3) ein, so erhält man bei der gewählten Schrittlänge für die Quadraturfehler  $R_\nu$  folgende Zusammenstellung:

**Tabelle 1**

|       | RV                   | SV                   | STV                  |
|-------|----------------------|----------------------|----------------------|
| $R_0$ | $17,2 \cdot 10^{-2}$ | $4,22 \cdot 10^{-3}$ | $6,56 \cdot 10^{-6}$ |
| $R_1$ | $9,4 \cdot 10^{-2}$  | $2,83 \cdot 10^{-3}$ | $4,36 \cdot 10^{-6}$ |

19. Soll nun auch der Einfluss der technischen Fehler berücksichtigt werden, so haben wir noch die entsprechenden Konstanten  $\omega$ ,  $\eta$  und  $\delta$  festzulegen. Wir setzen, indem wir an Stelle des Wertes von  $\omega_0$ , der sich aus (4) ergibt, eine halb so grosse individuelle Konstante  $\omega$  wählen,

$$\omega = \frac{5 \cdot 10^{-2}}{50} = 10^{-3}, \quad \eta = 2,5 \cdot 10^{-3}, \quad \delta = 3 \cdot 10^{-3}. \quad (26)$$

<sup>1)</sup> Durch eine sorgfältigere Abschätzung von (24) könnte man die Werte von  $m_1$  und  $m_0$  auf etwa die Hälfte herunterdrücken.

Für die Schranke  $\mu$  aus (15) kann, gemäss (21),  $\mu = 1,6$  gesetzt werden. Dann sind nach (16), (17) die Werte der Quadraturfehler  $R_0$  in Tabelle 1 in den Fällen des RV, SV, STV beziehungsweise um

$$3 \omega + 2 \eta = 8 \cdot 10^{-3}, \quad 5 \omega + 2 \eta = 10 \cdot 10^{-3}, \quad 4 \omega + 2 \eta = 9 \cdot 10^{-3}$$

zu vergrössern und die entsprechenden Werte von  $R_1$  zusätzlich noch um  $\delta \mu h = 8,4 \cdot 10^{-4}$ .

20. Führt man die Rechnungen aus und trägt alles in (22) ein, so erhält man schliesslich die folgende Zusammenstellung, in der die Schranken mit und ohne Berücksichtigung der technischen Fehler und ausserdem noch die entsprechenden Dimensionierungszahlen  $u$  aus (18) aufgeführt sind:

Tabelle 2

|                      | RV                           |      | SV                           |                      | STV                          |   |
|----------------------|------------------------------|------|------------------------------|----------------------|------------------------------|---|
|                      | $\omega = \eta = \delta = 0$ | (26) | $\omega = \eta = \delta = 0$ | (26)                 | $\omega = \eta = \delta = 0$ | - |
| $ \epsilon_9  \leq$  | 2,31                         | 2,45 | $6,53 \cdot 10^{-2}$         | $26,1 \cdot 10^{-2}$ | $10 \cdot 10^{-5}$           | — |
| $ \epsilon'_9  \leq$ | 1,23                         | 1,34 | $3,92 \cdot 10^{-2}$         | $19,0 \cdot 10^{-2}$ | $6 \cdot 10^{-5}$            | — |
| $\epsilon_9 =$       | -0,409                       |      | $1,6 \cdot 10^{-2}$          |                      | —                            |   |
| $\epsilon'_9 =$      | -0,182                       |      | $-0,72 \cdot 10^{-2}$        |                      | —                            |   |
| $u =$                | 0,052                        |      | 2,6                          |                      | $1,5 \cdot 10^3$             |   |

Man sieht, dass das RV typisch unterdimensioniert ist, und daher in diesem Falle der Schritt ruhig auf mehr als die Hälfte hätte verkleinert werden können. Das SV ist bereits leicht überdimensioniert, während das STV so stark überdimensioniert ist, dass es bei dieser Schrittweite überhaupt nicht durchzuführen ist.

Hingegen kann man sich fragen, um wieviel man (bei fester Zeicheneinheit  $e$ ) den Schritt vergrössern soll, damit auch das STV anwendbar wird. Nach dem Dimensionierungsprinzip des § 3 muss zu diesem Zweck  $u_{\text{STV}}$  in Tabelle 2 möglichst auf die Grössenordnung 1 gebracht werden. Begnügt man sich,  $u_{\text{STV}}$  ungefähr auf 10 herabzusetzen, so kann man, da  $u_{\text{STV}}$  ungefähr proportional  $10^{-5}$  ist,  $h$  mindestens um den Faktor 3 vergrössern.

Tatsächlich konnte bei dreifacher Schrittweite  $h = \pi/6$  das STV ohne Schwierigkeit durchgeführt werden. Man erhält das Ergebnis mit einem Fehler von  $\epsilon_3 = 1,8 \cdot 10^{-2}$ ,  $\epsilon'_3 = 0,68 \cdot 10^{-2}$ . Die entsprechenden Schranken lauten mit Einschluss der technischen Fehler  $9,1 \cdot 10^{-2}$  bzw.  $8,5 \cdot 10^{-2}$ . Die Schranken ohne Einschluss der technischen Fehler betragen  $0,83 \cdot 10^{-2}$  bzw.  $0,5 \cdot 10^{-2}$ , sind in diesem Fall also kleiner als die tatsächlich begangenen Fehler. Der genaue Wert von  $u$  beträgt  $u_{\text{STV}} = 10,8$ .

## Literaturverzeichnis

- BÖHMER, P. (1904): *Über geometrische Approximationen*, Diss. Göttingen.
- COLLATZ, L. (1942): *Natürliche Schrittweite bei numerischer Integration von Differentialgleichungssystemen*. Z. angew. Math. Mech. 22, 216–225.
- (1951): *Numerische Behandlung von Differentialgleichungen* (Springer, Berlin; Reihe: Grundlehren der mathematischen Wissenschaften, Bd. 60).
- GAUTSCHI, WALTER (1953): *Fehlerabschätzungen für die graphischen Integrationsverfahren von Grammel und Meissner-Ludwig*. Verh. Naturf. Ges. Basel 64, 401–435.
- GRAMMEL, R. (1939): *Ein Gegenstück zum Meissnerschen Verfahren der graphischen Analysis*. Ing.-Arch. 10, 395–411.
- KAMKE, E. (1944): *Differentialgleichungen, Lösungsmethoden und Lösungen*. Bd. 1, 3. Aufl. (Akademische Verlagsgesellschaft, Leipzig 1944).
- LUDWIG, R. (1951): *Graphische Integrationsverfahren für die Meissnerschen Linienbilder bei Differentialgleichungssystemen und Differentialgleichungen n-ter Ordnung*. Z. angew. Math. Mech. 31, 201–208.
- MEISSNER, E. (1931/32): *Graphische Analysis vermitteltst des Linienbildes einer Funktion*. Schweiz. Bauztg. 98, 287–290, 333–335 (1931); 99, 27–30, 41–44, 67–69 (1932). (Auch als Sonderabdruck im Verlag der Schweiz. Bauztg. 1932 erschienen.)
-

## Curriculum vitae

Als Sohn des Kaufmanns Hans Gautschi, wohnhaft in Basel, und der Margaritha geb. Eugster, wurde ich, WALTER GAUTSCHI, Bürger von Reinach (Kt. Aargau) und Basel, am 11. Dezember 1927 in Basel geboren. Nach Absolvierung der hiesigen Primarschule trat ich im Jahre 1938 in das Humanistische Gymnasium ein. Ein Jahr später wechselte ich in das Mathematisch-Naturwissenschaftliche Gymnasium und verliess dieses nach bestandener Maturität im Frühjahr 1947. Anschliessend bezog ich die hiesige Universität zum Studium der Mathematik, Physik, physikalischen Chemie und Versicherungsmathematik. 1951–1952 war ich Assistent bei Herrn Prof. Dr. A. Ostrowski.

Ich besuchte Vorlesungen und Praktika der folgenden Herren Professoren und Privatdozenten: E. Baldinger, E. Batschelet, P. Buchner, M. Fierz, P. Huber, W. Kuhn, E. Miescher, R. Nevanlinna, A. Ostrowski, E. von der Pahlen, E. Schubarth, A. Speiser, B. L. van der Waerden, E. Zwinggi.

Allen meinen Lehrern fühle ich mich zu grossem Dank verpflichtet. Besonderen Dank schulde ich Herrn Prof. Dr. A. Ostrowski, der meine Studien in besonderem Masse gefördert hat, und unter dessen Leitung die vorliegende Dissertation entstanden ist.





